# **PCT**





# INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classificati n <sup>5</sup> : A61K 39/02	A1	(11) International Publication Number: WO 94/21290 (43) International Publication Date: 29 September 1994 (29.09.94)
(21) International Application Number: PCT/US (22) International Filing Date: 15 March 1994 (		(81) Designated States: AU, BR, CA, FI, JP, KR, NO, RU, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB,
(30) Pri rity Data: 038,682 16 March 1993 (16.03.93)	τ	Published  S With international search report.
(71)(72) Applicants and Inventors: BARENKAMP, Ste [US/US]; 16 Villawood Lane, Webster Grove, MC 4954 (US). ST. GEME, Joseph, William, III [US. Bershire Drive, St. Louis, MO 63117 (US).	Ď 63119	) <u> </u>
(74) Agent: BERKSTRESSER, Jerry, W.; Shoemaker and Ltd., 2001 Jefferson Davis Highway, 1203 Cryst Building 1, P.O. Box 2286, Arlington, VA 222 (US).	tal Plaz	a
(54) Title: HIGH MOLECULAR WEIGHT SURFACE PR	ROTEI	IS OF NON-TYPEABLE HAEMOPHILUS

## (57) Abstract

High molecular weight surface proteins of non-typeable *Haemophilus influenzae* which exhibit immunogenic properties and genes encoding the same are described. Specifically, genes coding for two immunodominant high molecular weight proteins, HMW1 and HMW2, have been cloned, expressed and sequenced, while genes coding for high molecular proteins HMW3 and HMW4 have been cloned, expressed and partially sequenced.

#### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MIR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	ĦU	Hungary	NO	Norway
BG	Bulgaria	Œ	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic	SD	Sudan
CG	Congo		of Korea	SE	Sweden
CB	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
cs	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DĒ	Gennany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	MIL	Mali	UZ	Uzbekistan
FR	France	MIN	Mongolia	VN	Vict Nam
GA	Gahon		~		

WO 94/21290 PCT/US94/02550

#### TITLE OF INVENTION

# HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE HAEMOPHILUS

#### FIELD OF INVENTION

This invention relates to high molecular weight proteins of non-typeable haemophilus.

#### BACKGROUND TO THE INVENTION

Non-typeable <u>Haemophilus</u> <u>influenzae</u> are non-encapsulated organisms that are defined by their lack of reactivity with antisera against known <u>H. influenzae</u> capsular antigens.

commonly inhabit the upper organisms These tract of humans and are frequently respiratory infections, such as otitis media, responsible for sinusitis, conjunctivitis, bronchitis and pneumonia. Since these organisms do not have a polysaccharid are not controlled by the present capsule, they Haemophilus influenzae type b (Hib) vaccines, which are directed towards Hib bacterial capsular polysaccharides. The non-typeable strains, however, do produce surfac antigens that can elicit bactericidal antibodies. Two of the major outer membrane proteins, P2 and P6, have been identified as targets of human serum bactericidal activity. However, it has been shown that the P2 protein sequence is variable, in particular in the non-typeable Haemophilus strains. Thus, a P2-based vaccine would not protect against all strains of the organism.

There have previously been identified by Barenkamp et al (<u>Pediatr. Infect. Dis. J.</u>, 9:333-339, 1990) a group of high-molecular-weight (HMW) proteins that appeared to be major targets of antibodies present in human convalescent sera. Examination of a series of middle ar isolates revealed the presence of one or two such prot ins in m st strains. However, prior to the present inventi n, the structur s f these pr teins wer unknown as wer pure isolat s f such proteins.

5

10

15

20

25

30

35

10

15

20

25

30

### SUMMARY OF INVENTION

The inventors, in an effort to further characterize high molecular weight (HMW) Haemophilus proteins, have cloned, expressed and sequenced the genes coding for two immunodominant HMW proteins (designated HMW1 and HMW2) from a prototype non-typeable <u>Haemophilus</u> strain and have cloned, expressed and almost sequenced the genes coding for two additional immunodominant HMW proteins (designated HMW3 and HMW4) from another non-typeable Haemophilus strain.

In accordance with one aspect of the present invention, therefore, there is provided an isolated and purified gene coding for a high molecular weight protein of a non-typeable <u>Haemophilus</u> strain, particularly a gen coding for protein HMW1, HMW2, HMW3 or HMW4, as well as any variant or fragment of such protein which retains the immunological ability to protect against disease caused by a non-typeable <u>Haemophilus</u> strain. In another aspect, the invention provides a high molecular weight protein of non-typeable <u>Haemophilus</u> influenzae which is encoded by these genes.

#### BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is a DNA sequence of a gene coding for protein HMW1 (SEQ ID NO: 1);

Figure 2 is a derived amino acid sequence of protein HMW1 (SEQ ID NO: 2);

Figure 3 is a DNA sequence of a gene coding for protein HMW2 (SEQ ID NO: 3);

Figure 4 is a derived amino acid sequence of HMW2 (SEQ ID NO: 4);

Figure 5A shows restriction maps of representative recombinant phages which contained the HMW1 or HMW2 structural genes, the locations of the structural genes being indicated by the shaded bars;

Figure 5B sh ws th restricti n map f the T7 xpression vector pT7-7;

10

15

20

25

30

35

Figure 6 contains th DNA sequenc f a gen cluster for the <a href="https://mww.mww.mw.edu.nu...">https://mww.mw.edu.nu...</a> for the <a href="https://mw.edu.nu...">https://mw.edu.nu...</a> (SEQ ID NO: 5), comprising nucleotides 351 to 4958 (ORF a) (as in Figur 1), as well as two additional downstream genes in the 3' flanking region, comprising ORFs b, nucleotides 5114-6748 and c nucleotides 7062-9011:

Figure 7 contains the DNA sequence of a gene cluster for the <a href="https://mww.mw.edu.org/mw.edu

Figure 8 is a partial DNA sequence of a gene coding for protein HMW3 (SEQ ID NO: 7);

Figure 9 is a partial DNA sequence of a gene coding for protein HMW4 (SEQ ID NO: 8); and

Figure 10 is a comparison table for the derived amino acid sequence for proteins HMW1, HMW2, HMW3 and HMW4.

GENERAL DESCRIPTION OF INVENTION

The DNA sequences of the genes coding for HMW1 and HMW2, shown in Figures 1 and 3 respectively, were shown to be about 80% identical, with the first 1259 base pairs of the genes being identical. The derived amino acid sequences of the two HMW proteins, shown in Figures 2 and 4 respectively, are about 70% identical. Furthermore, the encoded proteins are antigenically related to the filamentous hemagglutinin surface protein of Bordetella A monoclonal antibody prepared against pertussis. filamentous hemagglutinin (FHA) of Bordetella pertussis was found to recognize both of the high molecular weight proteins. This data suggests that the HMW and FHA proteins may serve similar biological functions. Th derived amino acid sequences of the HMW1 and HMW2 pr teins sh w sequence similarity t that for the FHA protein. It has further been shown that this

antigenically-related proteins are produced by the majority of the non-typeabl strains of <u>Haemophilus</u>. Antisera raised against the protein express d by th <u>HMW1</u> gene recognizes both the <u>HMW2</u> protein and the <u>B. pertussis</u> FHA. The present invention includes an isolated and purified high molecular weight protein of non-typeable haemophilus which is antigenically related to the <u>B. pertussis</u> FHA, which may be obtained from natural sources or produced recombinantly.

10

15

5

A phage genomic library of a known strain of non-typeable <u>Haemophilus</u> was prepared by standard methods and the library was screened for clones expressing high molecular weight proteins, using a high titre antiserum against HMW's. A number of strongly reactive DNA clones were plaque-purified and sub-cloned into a T7 expression plasmid. It was found that they all expressed either on or the other of the two high-molecular-weight proteins designated HMW1 and HMW2, with apparent molecular weights of 125 and 120 kDa, respectively, encoded by open reading frames of 4.6 kb and 4.4 kb, respectively.

20

25

Representative clones expressing either HMW1 or HMW2 were further characterized and the genes isolated, purified and sequenced. The DNA sequence of HMW1 is shown in Figure 1 and the corresponding derived amino acid sequence in Figure 2. Similarly, the DNA sequence of HMW2 is shown in Figure 3 and the corresponding derived amino acid sequence in Figure 4. Partial purification of the isolated proteins and N-terminal sequence analysis indicated that the expressed proteins are truncated sinc their sequence starts at residue number 442 of both full length HMW1 and HMW2 gene products.

30

35

Subcloning studies with respect to the  $\underline{hmw1}$  and  $\underline{hmw2}$  genes indicated that correct processing of the HMW proteins required the products of additional downstream genes. It has been f und that both the  $\underline{hmw1}$  and  $\underline{hmw2}$  g nes ar flanked by two additional downstream p n

10

15

20

25

30

35

reading frames (ORFs), designated  $\underline{b}$  and  $\underline{c}$ , resp ctively, (see Figures 6 and 7).

The <u>b</u> ORFs are 1635 bp in length, extending from nucleotides 5114 to 6748 in the case of <u>hmwl</u> and nucleotides 5375 to 7009 in the case of <u>hmw2</u>, with their derived amino acid sequences 99% identical. The derived amino acid sequences demonstrate similarity with the derived amino acid sequences of two genes which encod proteins required for secretion and activation of hemolysins of <u>P. mirabilis</u> and <u>S. marcescens</u>.

The <u>c</u> ORFs are 1950 bp in length, extending from nucleotides 7062 to 9011 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mww.mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mww.mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case of <a href="https://mwl.and.nucleotides 7249">https://mwl.and.nucleotides 7249</a> to 9198 in the case

The two high molecular weight proteins have been isolated and purified and shown to be partially protective against otitis media in chinchillas and to function as adhesins. These results indicate th potential for use of such high molecular proteins and structurally-related proteins of other non-typeable strains of <u>Haemophilus influenzae</u> as components in non-typeable <u>Haemophilus influenzae</u> vaccines.

Since the proteins provided herein are good cross-reactive antigens and are present in the majority of non-typeable <u>Haemophilus</u> strains, it is evident that these HMW proteins may become integral constituents of a universal <u>Haemophilus</u> vaccine. Indeed, these proteins may be used not only as protective antigens against otitis, sinusitis and bronchitis caused by the non-typeable <u>Haemophilus</u> strains, but also may be used as carriers for the protectiv Hib polysaccharides in a c njugat vaccine against meningitis. The presents

10

15

20

25

30

35

may be used as carriers for other antig ns, haptens and polysaccharides from ther organisms, so as to induce immunity to such antigens, haptens and polysaccharides.

The nucleotide sequences encoding two high molecular weight proteins of a different non-typeable <u>Haemophilus</u> strain (designated HMW3 and HMW4) have been largely elucidated, and are presented in Figures 8 and 9. HMW3 has an apparent molecular weight of 125 kDa while HMW4 has an apparent molecular weight of 123 kDa. These high molecular weight proteins are antigenically related to the HMW1 and HMW2 proteins and to FHA. Sequence analysis of HMW3 is approximately 85% complete and of HMW4 95% complete, with short stretches at the 5'-ends of each gene remaining to be sequenced.

Figure 10 contains a multiple sequence comparison of the derived amino acid sequences for the four high molecular weight proteins identified herein. As may be seen from this comparison, stretches of identical peptide sequence may be found throughout the length of the comparison, with HMW3 more closely resembling HMW1 and HMW4 more closely resembling HMW2. This information is highly suggestive of a considerable sequence homology between high molecular weight proteins from various non-typeable <u>Haemophilus</u> strains.

In addition, mutants of non-typeable <u>H. influenzae</u> strains that are deficient in expression of HMW1 or HMW2 or both have been constructed and examined for their capacity to adhere to cultured human epithelial cells. The <a href="hmw1">hmw1</a> and <a href="hmw2">hmw2</a> gene clusters have been expressed in <u>E. coli</u> and have been examined for <a href="in vitro">in vitro</a> adherence. The results of such experimentation demonstrate that both HMW1 and HMW2 mediate attachment and hence are adhesins and that this function is present even in the absence of other <u>H. influenzae</u> surface structures.

With the isolati n and purificati n of the high mol cular w ight pr teins, th invent rs are abl to

10

15

20

25

30

determin th major pr t ctive epitopes by conventi nal epitope mapping and synth siz peptides c rresponding to thes d terminants to be incorporat d in fully synth ticor recombinant vaccines. Accordingly, the invention also comprises a synthetic peptide having an amino acid sequence corresponding to at least one protective epitope of a high molecular weight protein of a non-typeabl Haemophilus influenzae. Such peptides are of varying length that constitute portions of the molecular-weight proteins, that can be used to induc immunity, either directly or as part of a conjugate, against the relative organisms and thus constitut vaccines for protection against the corresponding diseases.

The present invention also provides any variant or fragment of the proteins that retains the potential immunological ability to protect against disease caused by non-typeable <u>Haemophilus</u> strains. The variants may be constructed by partial deletions or mutations of the genes and expression of the resulting modified genes to give the protein variations.

#### **EXAMPLES**

#### Example 1:

Non-typeable <u>H.influenzae</u> strains 5 and 12 were isolated in pure culture from the middle ear fluid of children with acute otitis media. Chromosomal DNA from strain 12, providing genes encoding proteins HMW1 and HMW2, was prepared by preparing Sau3A partial restriction digests of chromosomal DNA and fractionating on sucrose gradients. Fractions containing DNA fragments in the 9 to 20 kbp range were pooled and a library was prepared by ligation into  $\lambda$ EMBL3 arms. Ligation mixtures were packaged in vitro and plate-amplified in a P2 lysogen of E. coli LE392.

For plasmid subcl ning studi s, DNA from a representative recombinant phag was subcloned into th

10

15

20

25

30

35

T7 expr ssion plasmid pT7-7, containing the T7 RNA p lymerase promoter  $\Phi$ 10, a ribosom -binding sit and the translational start sit for th T7 gen 10 protein upstream from a multiple cloning site (see Figure 5B).

DNA sequence analysis was performed by the dideoxy method and both strands of the HMW1 gene and a single strand of the HMW2 gene were sequenced.

Western immunoblot analysis was performed identify the recombinant proteins being produced by reactive phage clones. Phage lysates grown in LE392 cells or plaques picked directly from a lawn of LE392 cells on YT plates were solubilized in qel electrophoresis sample buffer prior to electrophoresis. dodecvl sulfate (SDS)-polyacrylamide gel electrophoresis was performed on 7.5% 118 polyacrylamide modified Laemmli gels. After transfer of the proteins to nitrocellulose sheets, the sheets were probed sequentially with an E. coli-absorbed human serum sample containing high-titer antibody to the highmolecular-weight proteins and then with alkaline phosphatase-conjugated goat anti-human immunoglobulin G (IgG) second antibody. Sera from healthy adults contains high-titer antibody directed against surface-exposed high-molecular-weight proteins of non-typeable H. influenzae. One such serum sample was used as the screening antiserum after having been extensively absorbed with LE392 cells.

To identify recombinant proteins being produced by  $E.\ coli$  transformed with recombinant plasmids, the plasmids of interest were used to transform  $E.\ coli$  BL21 (DE3)/pLyss. The transformed strains were grown to an  $A_{600}$  of 0.5 in L broth containing 50  $\mu g$  of ampicillin per ml. IPTG was then added to 1 mM. One hour later, cells were harvested, and a sonicate of the cells was prepared. The protein concentrations of the samples were determined by the bicinchoninic acid method. Cell since

10

15

20

25

30

35

containing 100  $\mu$ g of t tal protein were s lubilized in el ctrophoresis sampl buffer, subjected to SDS-polyacrylamide gel el ctrophoresis, and transferred to nitrocellulose. The nitrocellulose was then probed sequentially with the <u>E. coli</u>-absorbed adult serum sampl and then with alkaline phosphatase-conjugated goat antihuman IgG second antibody.

Western immunoblot analysis also was performed to determine whether homologous and heterologous nontypeable H. influenzae strains expressed high-molecularweight proteins antigenically related to the protein encoded by the cloned HMW1 gene (rHMW1). Cell sonicates of bacterial cells were solubilized in electrophoresis sample buffer, subjected to SDS-polyacrylamide gel electrophoresis, and transferred to nitrocellulose. Nitrocellulose was probed sequentially with polyclonal rabbit rHMWl antiserum and then with alkaline phosphatase-conjugated goat anti-rabbit IgG second antibody.

Finally, Western immunoblot analysis was performed to determine whether non-typeable Haemophilus strains expressed proteins antigenically related to the filamentous hemagglutinin protein of Bordetella pertussis. Monoclonal antibody X3C, a immunoglobulin G antibody which (IgG) recognizes filamentous hemagglutinin, was used to probe cell sonicates by Western blot. An alkaline phosphataseconjugated goat anti-mouse IgG second antibody was used for detection.

To generate recombinant protein antiserum, <u>E. coli</u> BL21(DE3)/pLysS was transformed with pHMW1-4, and expression of recombinant protein was induced with IPTG, as described above. A cell sonicate of the bacterial cells was prepared and separated into a supernatant and pellet fraction by c ntrifugati n at 10,000 x g f r 30 min. The recombinant protein fraction nated with the

10

15

20

25

30

35

pell t fraction. A rabbit was subcutaneously immuniz d on biw ekly schedul with 1 mg f pr t in fr m th pell t fraction, the first dose given with Freund's compl t adjuvant and subsequent doses with Freund's incomplet adjuvant. Following the fourth injection, the rabbit was bled. Prior to use in the Western blot assay, the antiserum was absorbed extensively with sonicates of th host <u>E. coli</u> strain transformed with cloning vector alone.

To assess the sharing of antigenic determinants between HMW1 and filamentous hemagglutinin, enzyme-linked immunosorbent assay (ELISA) plates (Costar, Cambridge, Mass.) were coated with 60  $\mu$ l of a 4-ug/ml solution of filamentous hemagglutinin in Dulbecco's phosphatebuffered saline per well for 2 h at room temperature. Wells were blocked for 1 h with 1% bovine serum albumin in Dulbecco's phosphate-buffered saline prior to addition of serum dilutions. rHMW1 antiserum was serially diluted in 0.1% Brij (Sigma, St. Louis, Mo.) in Dulbecco's phosphate-buffered saline and incubated for 3 h at room temperature. After being washed, the plates were incubated with peroxidase-conjugated goat anti-rabbit lgG antibody (Bio-Rad) for 2 h at room temperature and subsedeveloped with 2,2'-azino-bis(3ethylbenzthiazoline-6-sulfonic acid) (Sigma) concentration of 0.54 in mg/ml in 0.1 M sodium citrate buffer, pH 4.2, containing 0.03% H<sub>2</sub>O<sub>2</sub>. Absorbances were read on an automated ELISA reader.

Recombinant phage expressing HMW1 or HMW2 were recovered as follows. The non-typeable <u>H. influenzae</u> strain 12 genomic library was screened for clones expressing high-molecular-weight proteins with an <u>E. coli</u>-absorbed human serum sample containing a high titer of antibodies directed against the high-molecular-weight proteins.

10

15

20

25

30

35

Num rous str ngly reactiv cl nes wer identified along with mor weakly reactiv nes. Twenty strongly clones were plaque-purified and examined by Western blot for expression of recombinant proteins. Each of the strongly reactive clones expressed one of two types of high-molecular-weight proteins, designated HMW1 and HMW2. The major immunoreactive protein bands in th HMW1 and HMW2 lysates migrated with apparent molecular masses of 125 and 120 kDa, respectively. In addition to the major bands, each lysate contained minor protein bands of higher apparent molecular weight. Protein bands seen in the HMW2 lysates at molecular masses of less than 120 kDa were not regularly observed and presumably represent proteolytic degradation products. Lysates of LE392 infected with the  $\lambda$ EMBL3 cloning vector alone were non-reactive when immunologically screened with the same serum sample. Thus, the observed activity was not due to cross-reactive E. coli proteins or AEMBL3-encoded pro-Furthermore, the recombinant proteins were not simply binding immunoglobulin nonspecifically, since the proteins were not reactive with the goat anti-human IgG conjugate alone, with normal rabbit sera, or with serum from a number of healthy young infants.

Representative clones expressing either the HMW1 or HMW2 recombinant proteins were characterized further. The restriction maps of the two phage types wer different from each other, including the regions encoding the HMW1 and HMW2 structural genes. Figure 5A shows restriction maps of representative recombinant phage which contained the HMW1 or HMW2 structural genes. Th locations of the structural genes are indicated by the shaded bars.

HMW1 plasmid subclones were constructed by using the T7 expression plasmid T7-7 (Fig. 5A and B). HMW2 plasmid subcl nes also were construct d, and the results with

th se latter subclon s were similar to those observed with the HMWl constructs.

The approximat location and direction transcription of the HMW1 structure gene were initially determined by using plasmid pHMW1 (Fig. 5A). plasmid was constructed by inserting the 8.5-kb BamHI-SalI fragment from \(\lambda\)HMW1 into \(\text{Bam}\)HI- and \(\text{Sal}\)I-cut pT7-7. E. coli transformed with pHMW1 expressed immunoreactive recombinant protein with an molecular mass of 115 kDa, which was strongly inducibl This protein was significantly smaller than with IPTG. the 125-kDa major protein expressed by the parent phage, indicating that it either was being expressed as a fusion protein or was truncated at the carboxy terminus.

To more precisely localize the 3' end of the structural gene, additional plasmids were constructed with progressive deletions from the 3' end of the pHMW1 construct. Plasmid pHMW1-1 was constructed by digestion of pHMW1 with PstI, isolation of the resulting 8.8-kb fragment, and religation. Plasmid pHMW1-2 constructed by digestion of pHMW1 with <u>HindIII</u>, isolation of the resulting 7.5-kb fragment, and religation. E. coli transformed with either plasmid pHMW1-1 or pHMW1-2 also expressed an immunoreactive recombinant protein with an apparent molecular mass of 115 kDa. These results indicated that the 3' end of the structural gene was 5' of the HindIII site.

To more precisely localize the 5' end of the gene, plasmids pHMW1-4 and pHMW1-7 were constructed. Plasmid pHMW1-4 was constructed by cloning the 5.1-kb <u>Bam</u>HI-<u>HindIII</u> fragment from \(\lambda\)HMW1 into a pT7-7-derived plasmid containing the upstream 3.8-kb <u>EcoRI-Bam</u>Hi fragment. <u>E. coli</u> transformed with pHMW1-4 expressed an immunoreactive protein with an apparent molecular mass of approximately 160 kDa. Although pr tein production was inducible with IPTG, the local value of the production in these

5

10

15

20

25

30

35

10

15

20

25

30

35

transformants were substantially lower than those with the pHMW1-2 transformants described above. pHMW1-7 was constructed by digesting pHMW1-4 with NdeI and SpeI. The 9.0-kbp fragment generated by this double digestion was isolated, blunt ended, and religated. coli transformed with pHMW1-7 also expressed an immunoreactive protein with an apparent molecular mass of 160 kDa, a protein identical in size to that expressed by the pHMW1-4 transformants. The result indicated that the initiation codon for the HMW1 structural gene was 3' of the SpeI site. DNA sequence analysis confirmed this conclusion.

As noted above, the \(\lambda\)HMW1 phage clones expressed a major immunoreactive band of 125 kDa, whereas the HMW1 plasmid clones pHMW1-4 and pHMW1-7, which contained what was believed to be the full-length gene, expressed an immunoreactive protein of approximately 160 kDa. size discrepancy was disconcerting. One possible explanation was that an additional gene or necessary for correct processing of the HMW1 gene product were deleted in the process of subcloning. To address this possibility, plasmid pHMW1-14 was constructed. This construct was generated by digesting pHMW1 with NdeI and MluI and inserting the 7.6-kbp NdeI-MluI fragment isolated from pHMW1-4. Such a construct would contain the full-length HMW1 gene as well as the DNA 3' of th HMW1 gene which was present in the original HMW1 phage. E. coli transformed with this plasmid expressed major immunoreactive proteins with apparent molecular masses of and 160 kDa as well as additional degradation The 125- and 160-kDa bands were identical t the major and minor immunoreactive bands detected in th HMWl phage lysates. Interestingly, the pHMW1-14 construct also expressed significant amounts of protein in th uninduced condition, a situati n not bserved with th earlier c nstructs.

The relationship b tw en the 125- and 160-kDa proteins remains somewhat unclear. S quence analysis, described b low, r v als that the HMW1 gen would be predicted to encode a protein of 159 kDa. It is believed that the 160-kDa protein is a precursor form of the mature 125-kDa protein, with the conversion from one protein to the other being dependent on the products of the two downstream genes.

Sequence analysis of the HMW1 gene (Figure 1) revealed a 4,608-bp open reading frame (ORF), beginning with an ATG codon at nucleotide 351 and ending with a TAG stop codon at nucleotide 4959. A putative ribosomebinding site with the sequence AGGAG begins 10 bp upstream of the putative initiation codon. Five other inframe ATG codons are located within 250 bp of th beginning of the ORF, but none of these is preceded by a typical ribosome-binding site. The 5'-flanking region of the ORF contains a series of direct tandem repeats, with the 7-bp sequence ATCTTTC repeated 16 times. tandem repeats stop 100 bp 5' of the putative initiation codon. An 8-bp inverted repeat characteristic of a rhoindependent transcriptional terminator is present, beginning at nucleotide 4983, 25 bp 3' of the presumed translational stop. Multiple termination codons are present in all three reading frames both upstream and downstream of the ORF. The derived amino acid sequenc of the protein encoded by the HMW1 gene (Figure 2) has a molecular weight of 159,000, in good agreement with the apparent molecular weights of the proteins expressed by the HMW1-4 and HMW1-7 transformants. The derived amino acid sequence of the amino terminus does not demonstrate the characteristics of a typical signal sequence. BamHI site used in generation of pHMW1 comprises bp 1743 through 1748 of the nucleotide sequence. d wnstream of the BamHI site would be predicted to encod a pr tein of 111 kDa, in good agreem nt with th 115 kDa

5

10

15

20

25

30

35

10

15

20

25

30

35

estimated for the apparent molecular mass f the pHMW1-encoded fusion protein.

The s quence of the HMW2 gene (Figure 3) consists of a 4,431-bp ORF, beginning with an ATG codon at nucleotide 352 and ending with a TAG stop codon at nucleotide 4783. The first 1,259 bp of the ORF of the HMW2 gene ar identical to those of the HMW1 gene. Thereafter, the sequences begin to diverge but are 80% identical overall. With the exception of a single base addition at nucleotide 93 of the HMW2 sequence, the 5'-flanking regions of the HMW1 and HMW2 genes are identical for 310 bp upstream from the respective initiation codons. Thus, the HMW2 gene is preceded by the same set of tandem repeats and the same putative ribosome-binding site which lies 5' of the HMW1 gene. A putative transcriptional terminator identical to that identified 3' of the HMW1 is noted, beginning at nucleotide 4804. The discrepancy in the lengths of the two genes principally accounted for by a 186-bp gap in the HMW2 sequence, beginning at nucleotide position 3839. derived amino acid sequence of the protein encoded by th HMW2 gene (Figure 4) has a molecular weight of 155,000 and is 71% identical with the derived amino acid sequenc of the HMW1 gene.

The derived amino acid sequences of both the HMW1 and HMW2 genes (Figures 2 and 4) demonstrated sequence similarity with the derived amino acid sequence of filamentous hemagglutinin of Bordetella pertussis, a surface-associated protein of this organism. The initial and optimized TFASTA scores for the HMW1-filamentous hemagglutinin sequence comparison were 87 and 186, respectively, with a word size of 2. The z score for the comparison was 45.8. The initial and optimized TFASTA scores for the HMW2-filamentous hemagglutinin sequence comparison were 68 and 196, respectively. The z score for the latter comparison was 48.7. The magnitudes of

th initial and optimized TFASTA scores and the z scores suggest d that a biologically significant relati nship existed between the HMW1 and HMW2 gene products and filamentous hemagglutinin. When the derived amino acid sequences of HMW1, HMW2, and filamentous hemagglutinin genes were aligned and compared, the similarities were most notable at the amino-terminal ends of the three sequences. Twelve of the first 22 amino acids in the predicted peptide sequences were identical. additional, the sequences demonstrated a common fiveamino-acid stretch, Asn-Pro-Asn-Gly-Ile, and several shorter stretches of sequence identity within the first 200 amino acids.

#### Example 2:

5

10

15

20

25

30

35

further To explore the HMW1-filamentous hemagglutinin relationship, the ability of antiserum prepared against the HMW1-4 recombinant protein (rHMW1) to recognize purified filamentous hemagglutinin was The rHMW1 antiserum demonstrated ELISA assessed. reactivity with filamentous hemagglutinin in a dosedependent manner. Preimmune rabbit serum had minimal reactivity in this assay. The rHMW1 antiserum also was examined in a Western blot assay and demonstrated weak positive reactivity with purified filamentous hemagglutinin in this system also.

To identify the native <u>Haemophilus</u> protein corresponding to the HMWl gene product and to determine the extent to which proteins antigenically related to the HMWl cloned gene product were common among other non-typeable <u>H. influenzae</u> strains, a panel of <u>Haemophilus</u> strains was screened by Western blot with the rHMWl antiserum. The antiserum recognized both a 125- and a 120-kDa protein band in the homologous strain 12, the putative mature protein products of the HMWl and HMW2 genes, resp ctively.

25

30

35

When used to scr en heterol gous n n-typ abl influenzae strains, rHMW1 antis rum recognized highm lecular-weight proteins in 75% of 125 epidemiologically In general, the antiserum reacted unrelated strains. with one or two protein bands in the 100- to 150-kDa range in each of the heterologous strains in a pattern similar but not identical to that seen in the homologous strain.

Monoclonal antibody X3C is a murine IgG antibody directed against the filamentous hemagglutinin protein of 10 B. pertussis. This antibody can inhibit the binding of B. pertussis cells to Chinese hamster ovary cells and HeLa cells in culture and will inhibit hemagglutination of erythrocytes by purified filamentous hemagglutinin. 15 A Western blot assay was performed in which this monoclonal antibody was screened against the same panel of non-typeable H. influenzae strains discussed above. Monoclonal antibody X3C recognized both the highmolecular-weight proteins in non-typeable H. influenza 20 strain 12 which were recognized by the recombinantprotein antiserum. In addition, the monoclonal antibody recognized protein bands in a subset of heterologous nontypeable H. influenzae strains which were identical to those recognized by the recombinant-protein antiserum. On occasion, the filamentous hemagglutinin monoclonal antibody appeared to recognize only one of the two bands which had been recognized by the recombinant-protein antiserum. Overall, monoclonal antibody X3C recognized high-molecular-weight protein bands identical to thos recognized by the rHMW1 antiserum in approximately 35% of our collection of non-typeable H. influenzae strains. Example 3:

Mutants deficient in expression of HMW1, MW2 or both proteins were constructed to examine the role of these proteins in bacterial adh r nce. Th following strategy was employed. pHMW1-14 (see Example 1, Figure 5A) was

digested with BamHI and then ligated to a kanamycin cassette isolated on a 1.3-kb BamHl fragment from pUC4K. r sultant plasmid (pHMW1-17) was lin arized by digestion with XbaI and transformed into non-typeable H. influenzae strain 12, followed by selection for kanamycin Southern analysis of a series of resistant colonies. these colonies demonstrated two populations transformants, one with an insertion in the HMW1 structural gene and the other with an insertion in the HMW2 structural gene. One mutant from each of these classes was selected for further studies.

Mutants deficient in expression of both proteins were recovered using the following protocol. deletion of the 2.1-kb fragment of DNA between two EcoRI sites spanning the 3'-portion of the HMW1 structural gene in pHMW-15, the kanamycin cassette from pUC4K was inserted as a 1.3-kb EcoRl fragment. The resulting plasmid (pHMW1-16) was linearized by digestion with XbaI and transformed into strain 12, followed again by selection for kanamycin resistant colonies. Southern analysis of a representative sampling of these colonies demonstrated that in seven of eight cases, insertion into both the HMW1 and HMW2 loci had occurred. One such mutant was selected for further studies.

To confirm the intended phenotypes, the mutant strains were examined by Western blot analysis with a polyclonal antiserum against recombinant HMW1 protein. The parental strain expressed both the 125-kD HMW1 and the 120-kD HMW2 protein. In contrast, the HMW2 mutant failed to express the 120-kD protein, and the HMW1 mutant failed to express the 125-kD protein. The double mutant lacked expression of either protein. On the basis of whole cell lysates, outer membrane profiles, and colony morphology, the wild type strain and the mutants were otherwise identical with ne anoth r. Transmissi n

5

10

15

20

25

30

35

10

15

20

25

30

35

electron microscopy d monstrated that none of the four strains expressed pili.

The capacity of wild type strain 12 to adhere to Chang epithelial cells was examined. In such assays, bacteria were inoculated into broth and allowed to grow to a density of ~2 x  $10^9$  cfu/ml. Approximately 2 x  $10^7$ cfu were inoculated onto epithelial cell monolayers, and plates were gently centrifuged at 165 x g for 5 minutes to facilitate contact between bacteria and the epithelial surface. After incubation for 30 minutes at 37°C in 5% CO,, monolayers were rinsed 5 times with PBS to remove nonadherent organisms and were treated with trypsin-EDTA (0.05% trypsin, 0.5% EDTA) in PBS to release them from the plastic support. Well contents were agitated, and dilutions were plated on solid medium to yield the number of adherent bacteria per monolayer. Percent adherenc was calculated by dividing the number of adherent cfu per monolayer by the number of inoculated cfu.

As depicted in Table 1 below (the Tables appear at the end of the descriptive text), this strain adhered quite efficiently, with nearly 90% of the inoculum binding to the monolayer. Adherence by the mutant expressing HMW1 but not HMW2 (HMW2') was also quite efficient and comparable to that by the wild type strain. In contrast, attachment by the strain expressing HMW2 but deficient in expression of HMW1 (HMW1') was decreased about 15-fold relative to the wild type. Adherence by the double mutant (HMW1'/HMW2') was decreased even further, approximately 50-fold compared with the wild type and approximately 3-fold compared with the HMW1 mutant. Considered together, these results suggest that both the HMW1 protein and the, HMW2 protein influence attachment to Chang epithelial cells. Interestingly, optimal adherence to this cell line appears to require HMW1 but not HMW2.

10

15 '

20

25

30

35

### Example 4:

plasmids pHMW1-16 and pHMW1-17 Using th Example 3) and following a scheme similar to that employed with strain 12 as described in Example 3, three non-typeable Haemophilus strain 5 mutants were isolated, including one with the kanamycin gene inserted into the hmwl-like (designated hmw3) locus, a second with an insertion in the <a href="https://mww.energia.com/">https://mww.energia.com/</a> locus, and a third with insertions in both loci. As predicted, Western immunoblot analysis demonstrated that the mutant with insertion of the kanamycin cassette into the hmwllike locus had lost expression of the HMW3 125-kD protein, while the mutant with insertion into the hmw2like locus failed to express the HMW4 123-kD protein. The mutant with a double insertion was unable to express either of the high molecular weight proteins.

As shown in Table 1 below, wild type strain 5 demonstrated high level adherence, with almost 80% of the inoculum adhering per monolayer. Adherence by the mutant deficient in expression of the HMW2-like protein was also quite high. In contrast, adherence by the mutant unabl to express the, HMW1-like protein was reduced about 5-fold relative to the wild type, and attachment by the double mutant was diminished even further (approximately 25-fold). Examination of Giemsa-stained samples confirmed these observations (not shown). Thus, the results with strain 5 corroborate the findings with strain 12 and the HMW1 and HMW2 proteins.

#### Example 5:

To confirm an adherence function for the HMW1 and HMW2 proteins and to examine the effect of HMW1 and HMW2 independently of other <u>H. influenzae</u> surface structures, the <u>hmw1</u> and the <u>hmw2</u> gene clusters were introduced into <u>E. coli</u> DH5\alpha, using plasmids pHMW1-14 and pHMW2-21, respectively. As a control, the claim vect r, pT7-7, was also transformed into <u>E. coli</u> DH5\alpha. Western blot

10

15

20

25

30

35

Example 6:

analysis demonstrated that <u>E. coli</u> DH5 $\alpha$  containing the <u>hmwl</u> g nes express d a 125 kDa protein, whil the same strain harboring the <u>hmw2</u> genes expressed a 120-kDa protein. <u>E. coli</u> DH5 $\alpha$  containing pT7-7 failed to react with antiserum against recombinant HMW1. Transmission electron microscopy revealed no pili or other surfac appendages on any of the <u>E. coli</u> strains.

Adherence by the E. coli strains was quantitated and compared with adherence by wild type non-typeable H. influenzae strain 12. As shown in Table 2 below, adherence by E. coli DH5 $\alpha$  containing vector alone was less than 1% of that for strain 12. In contrast, E. coli DH5α harboring the <u>hmwl</u> gene cluster demonstrated adherence levels comparable to those for strain 12. Adherence by E. coli DH5a containing the hmw2 genes was approximately 6-fold lower than attachment by strain 12 but was increased 20-fold over adherence by E. coli DH5a with pT7-7 alone. These results indicate that the HMW1 and HMW2 proteins are capable of independently mediating attachment to Chang conjunctival cells. These results are consistent with the results with the H. influenzae mutants reported in Examples 3 and 4, providing further evidence that, with Chang epithelial cells, HMW1 is a more efficient adhesin than is HMW2.

Experiments with <u>E. coli</u> HB101 harboring pT7-7, pHMW1-14, or pHMW2-21 confirmed the results obtained with the DH5 $\alpha$  derivatives (see Table 2).

HMW1 and HMW2 were isolated and purified from non-typeable <u>H. influenzae</u> (NTHI) strain 12 in the following manner. Non-typeable <u>Haemophilus</u> bacteria from frozen stock culture were streaked onto a chocolate plate and grown overnight at 37°C in an incubator with 5%  $\rm CO_2$ . 50ml starter culture of brain heart infusion (BHI) broth, supplement d with 10  $\mu$ g/ml each f hemin and NAD was inoculated with growth on chocolat plate. The start r

10

15

20

25

30

35

culture was grown until the optical density (0.D. - 600nm) reached 0.6 to 0.8 and then the bacteria in the starter culture was used to inoculat six 500 ml flasks of supplemented BHI using 8 to 10 ml per flask. The bacteria were grown in 500 ml flasks for an additional 5 to 6 hours at which time the 0.D. was 1.5 or greater. Cultures were centrifuged at 10,000 rpm for 10 minutes.

Bacterial pellets were resuspended in a total volume of 250 ml of an extraction solution comprising 0.5 M NaCl. 0.01 M Na,EDTA, 0.01 M Tris 50  $\mu$ M phenanthroline, pH 7.5. The cells were not sonicated or otherwise disrupted. The resuspended cells were allowed to sit on ice at 0°C for 60 minutes. The resuspended cells were centrifuged at 10,000 rpm for 10 minutes at 4°C to remove the majority of intact cells and cellular debris. The supernatant was collected and centrifuged at 100,000 xg for 60 minutes at 4°C. The supernatant again was collected and dialyzed overnight at 4°C against 0.01 M sodium phosphate, pH 6.0.

The sample was centrifuged at 10,000 rpm for 10 minutes at 4°C to remove insoluble debris precipitated from solution during dialysis. The supernatant was applied to a 10 ml CM Sepharose column which has been pre-equilibrated with 0.01 M sodium phosphate, pH 6. Following application to this column, the column was washed with 0.01 M sodium phosphate. Proteins were elevated from the column with a 0 - 0.5M KCl gradient in 0.01 M Na phosphate, pH 6 and fractions were collected for gel examination. Coomassie gels of column fractions were carried out to identify those fractions containing high molecular weight proteins. The fractions containing high molecular weight proteins were pooled concentrated to a 1 to 3 ml volume in preparation for application of sample to gel filtration column.

A Sepharose CL-4B gel filtration column was equilibrated with phosphat -buff red saline, pH 7.5. The

15

20

25

30

35

Ĺ

'*द* 

1

, ...

. 4

concentrated high molecular weight protein sample was applied to the gel filtration column and column fractions w re collect d. Coomassi g ls w re performed on the column fractions to identify those containing high molecular weight proteins. The column fractions containing high molecular weight proteins were pooled.

The proteins were tested to determine whether they would protect against experimental otitis media caused by the homologous strain.

10 Chinchillas received three monthly subcutaneous injections with 40  $\mu$ g of an HMW1-HMW2 protein mixture in Freund's adjuvant. One month after the last injection, the animals were challenged by intrabullar inoculation with 300 cfu of NTHI strain 12.

Infection developed in 5 of 5 control animals versus 5 of 10 immunized animals. Among infected animals, geometric mean bacterial counts in middle ear fluid 7 days post-challenge were  $7.4 \times 10^6$  in control animals verus  $1.3 \times 10^5$  in immunized animals.

Serum antibody titres following immunization were comparable in uninfected and infected animals. However, infection in immunized animals was uniformly associated with the appearance of bacteria down-regulated in expression of the HMW proteins, suggesting bacterial selection in response to immunologic pressure.

Although this data shows that protection following immunization was not complete, this data suggests the HMW adhesin proteins are potentially important protective antigens which may comprise one component of a multicomponent NTHI vaccine.

These animal challenge tests were repeated in Chinchillas at a lower dose challenge than the 300 cfu employed above. In this instance, complete protection was achieved. In these experiments, groups of five animals were immunized with 20  $\mu$ g of th HMW1-HMW2

mixture on days 1, 28, and 42 in the presence of AlPO<sub>4</sub>. Blood sampl s wer coll ct d on day 53 to monitor the antibody response. On day 56, the 1 ft ear of animals was challenged with about 10 cfu of <u>H. influenzae</u> strain 12. Ear infection was monitored on day 4. Four animals in Group 3 were infected previously by <u>H. influenzae</u> strain 12 and were recovered completely for at least on month before the second challenge. The results are outlined in the following Table A:

10 TABLE A

Protective ability of HMW protein against non-typeable <u>H. influenzae</u> challenge in chinchilla model

15

20

Group	Antigens	Total Animals	Number of Animals Showed Positive Ear Infection			
(#)			Tympano- gram	Otosco- pic Examin- ation	cfu of Bac- teria/ 10 µL	
1	HMW	5	0	0	0	
2	None	5	5	5	850- 3200 (4/5)	
3	Convalescent	4	0	0	0	

25

30

35

# Example 7:

A number of synthetic peptides were derived from HMW1. Antisera then was raised to these peptides. anti-peptide antisera to peptide HMW1-P5 was shown to recognize HMW1. Peptide HMW1-P5 covers amino acids 1453 to 1481 of HMW1, has the sequence **VDEVIEAKRILEKVKDLSDEEREALAKLG** (SEO ID NO:9), and represents bases 1498 to 1576 in Figure 10.

This finding d monstrates that the DNA sequ nce and the d rived protein is being interpreted in the c rrect

reading frame and that p ptides deriv d from the sequence can be produced which will be immunogenic.

# SUMMARY OF DISCLOSURE

In summary of this disclosure, the present invention provides high molecular weight proteins of non-typeable Haemophilus, genes coding for the same and vaccines incorporating such proteins. Modifications are possibl within the scope of this invention.

Table 1. Effect of mutation of high m lecular weight proteins on adherence to Chang epithelial cells by nontypable H. influenzae.

# ADHERENCE\*

Strain	jnoculum j	relative to wild type†				
Strain 12 derivatives						
wild type	87.7 ± 5.9	$100.0 \pm 6.7$				
HMW1-mutant	6.0 <u>+</u> 0.9	6.8 ± 1.0				
HMW2-mutant	89.9 ± 10.8	$102.5 \pm 12.3$				
HMW1-/HMW2- mutant	$2.0 \pm 0.3$ $2.3 \pm 0.3$					
Strain 5 derivatives						
wild type	$78.7 \pm 3.2$	$100.0 \pm 4.1$				
HMW1-like mutant	$15.7 \pm 2.6$	$19.9 \pm 3.3$				
HMW2-like mutant	$103.7 \pm 14.0$	$131.7 \pm 17.8$				
double mutant	$3.5 \pm 0.6$	4.4 <u>+</u> 0.8				
		•				

<sup>\*</sup>Numbers represent mean (+ standard error of the mean) of measurements in triplicate or quadruplicate from representative experiments.

<sup>†</sup> Adherence values for strain 12 derivatives are relative to strain 12 wild type; values for strain 5 derivatives are relative to strain 5 wild type.

Table 2. Adherence by E. coli DH5a and HB101 harboring hmwl or hmw2 gene clusters.

	Adherence relative to
Strain*	H. influenzae strain 12†
DH5α (pT7-7)	$0.7 \pm 0.02$
DH5α (pHMW1-14)	114.2 ± 15.9
DH5α (pHMW2-21)	$14.0 \pm 3.7$
HB101 (pT7-7)	1.2 ± 0.5
HB101 (pHMW1-14)	93.6 ± 15.8
HB101 (pHMW2-21)	3.6 <u>+</u> 0.9

The plasmid pHMW1-14 contains the *hmwl* gene cluster, while pHMW2-21 contains the *hmw2* gene cluster; pT7-7 is the cloning vector used in these constructs.

<sup>†</sup> Numbers represent the mean (+ standard error of the mean) of measurements made in triplicate from representative experiments.

#### SEQUENCE LISTING

(1) GENERAL	INFORMATION:
-------------	--------------

- (i) APPLICANT: BARENKAMP, STEPHEN J ST. GEME III, JOSEPH W
- (ii) TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE HAEMOPHILUS
- (iii) NUMBER OF SEQUENCES: 8
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Shoemaker and Mattare, Ltd
  - (B) STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza Bldg. 1
  - (C) CITY: Arlington
  - (D) STATE: Virginia
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 22202-0286
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk

  - (B) COMPUTER: IBM PC compatible (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/038,682
  - (B) FILING DATE: 16-MAR-1993
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: BERKSTRESSER, JERRY W
  - (B) REGISTRATION NUMBER: 22,651
  - (C) REFERENCE/DOCKET NUMBER: 1038-293
  - (ix) TELECOMMUNICATION INFORMATION:
    - (A) TELEPHONE: (703) 415-0810
    - (B) TELEFAX: (703) 415-0813
- (2) INFORMATION FOR SEQ ID NO:1:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 5116 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACAGCGTTCT	CTTAATACTA	GTACAAACCC	ACAATAAAAT	ATGACAAACA	ACAATTACAA	60
CACCTTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAATA	GTATAAATCC	GCCATATAAA	120
ATGGTATAAT	CTTTCATCTT	TCATCTTTCA	TCTTTCATCT	TTCATCTTTC	ATCTTTCATC	180
TTTCATCTTT	CATCTTTCAT	CTTTCATCTT	TCATCTTTCA	TCTTTCATCT	TTCATCTTTC	240
ACATGCCCTG	ATGAACCGAG	GGAAGGGAGG	GAGGGGCAAG	AATGAAGAGG	GAGCTGAACG	300

SUBSTITUTE SHEET (RULE 26)

AACGCAAATG	ATAAAGTAA1	TTAATTGTTC	AACTAACCTT	AGGAGAAAAT	ATGAACAAGC	360
TATATCGTCT	CAAATTCAGO	AAACGCCTGA	ATGCTTTGGT	TGCTGTGTCT	GAATTGGCAC	420
GGGGTTGTGA	CCATTCCACA	GAAAAAGGCA	GCGAAAAACC	TGCTCGCATG	AAAGTGCGTC	480
ACTTAGCGTT	AAAGCCACTI	TCCGCTATGT	TACTATCTTT	AGGTGTAACA	TCTATTCCAC	540
AATCTGTTTI	AGCAAGCGGC	TTACAAGGAA	TGGATGTAGT	ACACGGCACA	GCCACTATGC	600
AAGTAGATGG	TAATAAAACC	ATTATCCGCA	ACAGTGTTGA	CGATATCATT	AATTGGAAAC	660
AATTTAACAT	CGACCAAAAT	GAAATGGTGC	AGTTTTTACA	AGAAAACAAC	AACTCCGCCG	720
TATTCAACCG	TGTTACATCT	AACCAAATCT	CCCAATTAAA	AGGGATTTTA	GATTCTAACG	780
GACAAGTCTT	TTTAATCAAC	CCAAATGGTA	TCACAATAGG	TAAAGACGCA	ATTATTAACA	840
CTAATGGCTT	TACGGCTTCT	ACGCTAGACA	TTTCTAACGA	AAACATCAAG	GCGCGTAATT	900
TCACCTTCGA	GCAAACCAAA	GATAAAGCGC	TCGCTGAAAT	TGTGAATCAC	GGTTTAATTA	960
CTGTCGGTAA	AGACGGCAGT	GTAAATCTTA	TTGGTGGCAA	AGTGAAAAAC	GAGGGTGTGA	1020
TTAGCGTAAA	TGGTGGCAGC	ATTTCTTTAC	TCGCAGGGCA	AAAAATCACC	ATCAGCGATA	1080
TAATAAACCC	AACCATTACT	TACAGCATTG	CCGCGCCTGA	AAATGAAGCG	GTCAATCTGG	1140
GCGATATTTT	TGCCAAAGGC	GGTAACATTA	ATGTCCGTGC	TGCCACTATT	CGAAACCAAG	1200
GTAAACTTTC	TGCTGATTCT	GTAAGCAAAG	ATAAAAGCGG	CAATATTGTT	CTTTCCGCCA	1260
AAGAGGGTGA	AGCGGAAATT	GGCGGTGTAA	TTTCCGCTCA	AAATCAGCAA	GCTAAAGGCG	1320
GCAAGCTGAT	GATTACAGGC	GATAAAGTCA	CATTAAAAAC	AGGTGCAGTT	ATCGACCTTT	1380
CAGGTAAAGA	AGGGGGAGAA	ACTTACCTTG	GCGGTGACGA	GCGCGGCGAA	GGTAAAAAGG	1440
GCATTCAATT	AGCAAAGAAA	ACCTCTTTAG	AAAAAGGCTC	AACCATCAAT	GTATCAGGCA	1500
AAGAAAAAGG	CGGACGCGCT	ATTGTGTGGG	GCGATATTGC	GTTAATTGAC	GGCAATATTA	1560
ACGCTCAAGG	TAGTGGTGAT	ATCGCTAAAA	CCGGTGGTTT	TGTGGAGACG	TCGGGGCATG	1620
ATTTATTCAT	CAAAGACAAT	GCAATTGTTG	ACGCCAAAGA	GTGGTTGTTA	GACCCGGATA	1680
ATGTATCTAT	TAATGCAGAA	ACAGCAGGAC	GCAGCAATAC	TTCAGAAGAC	GATGAATACA	1740
CGGGATCCGG	GAATAGTGCC	AGCACCCCAA	AACGAAACAA	AGAAAAGACA	ACATTAACAA	1800
ACACAACTCT	TGAGAGTATA	CTAAAAAAAG	GTACCTTTGT	TAACATCACT	GCTAATCAAC	1860
GCATCTATGT	CAATAGCTCC	ATTAATTTAT	CCAATGGCAG	CTTAACTCTT	TGGAGTGAGG	1920
GTCGGAGCGG	TGGCGGCGTT	GAGATTAACA	ACGATATTAC	CACCGGTGAT	GATACCAGAG	1980
GTGCAAACTT	AACAATTTAC	TCAGGCGGCT	GGGTTGATGT	TCATAAAAAT	ATCTCACTCG	2040
GGGCGCAAGG	TAACATAAAC	ATTACAGCTA	AACAAGATAT	CGCCTTTGAG	aaaggaagca	2100
ACCAAGTCAT	TACAGGTCAA	GGGACTATTA	CCTCAGGCAA	TCAAAAAGGT	TTTAGATTTA	2160
ATAATGTCTC	TCTAAACGGC	ACTGGCAGCG	GACTGCAATT	CACCACTAAA	AGAACCAATA	2220
AATACGCTAT	CACAAATAAA	TTTGAAGGGA	CTTTAAATAT	TTCAGGGAAA	GTGAACATCT	2280
CAATGGTTTT	ACCTAAAAAT	GAAAGTGGAT	ATGATAAATT	CAAAGGACGC	ACTTACTGGA	2340

ATTTAACCTC	CTTAAATGTT	TCCGAGAGTG	GCGAGTTTAA	CCTCACTATT	GACTCCAGAG	240
GAAGCGATAG	TGCAGGCACA	CTTACCCAGC	CTTATAATTT	AAACGGTATA	TCATTCAACA	246
AAGACACTAC	CTTTAATGTT	GAACGAAATG	CAAGAGTCAA	CTTTGACATC	AAGGCACCAA	252
TAGGGATAAA	TAAGTATTCT	AGTTTGAATT	ACGCATCATT	TAATGGAAAC	ATTTCAGTTT	258
CGGGAGGGG	GAGTGTTGAT	TTCACACTTC	TCGCCTCATC	CTCTAACGTC	CAAACCCCCG	264
GTGTAGTTAT	AAATTCTAAA	TACTTTAATG	TTTCAACAGG	GTCAAGTTTA	AGATTTAAAA	270
CTTCAGGCTC	AACAAAAACT	GGCTTCTCAA	TAGAGAAAGA	TTTAACTTTA	AATGCCACCG	276
GAGGCAACAT	AACACTTTTG	CAAGTTGAAG	GCACCGATGG	AATGATTGGT	AAAGGCATTG	2820
TAGCCAAAAA	AAACATAACC	TTTGAAGGAG	GTAACATCAC	CTTTGGCTCC	AGGAAAGCCG	2880
TAACAGAAAT	CGAAGGCAAT	GTTACTATCA	ATAACAACGC	TAACGTCACT	CTTATCGGTT	2940
CGGATTTTGA	CAACCATCAA	AAACCTTTAA	СТАТТААААА	AGATGTCATC	ATTAATAGCG	3000
GCAACCTTAC	CGCTGGAGGC	AATATTGTCA	ATATAGCCGG	AAATCTTACC	GTTGAAAGTA	3060
ACGCTAATTT	CAAAGCTATC	ACAAATTTCA	CTTTTAATGT	AGGCGGCTTG	TTTGACAACA	3120
AAGGCAATTC	AAATATTTCC	ATTGCCAAAG	GAGGGGCTCG	CTTTAAAGAC	ATTGATAATT	3180
CCAAGAATTT	AAGCATCACC	ACCAACTCCA	GCTCCACTTA	CCGCACTATT	ATAAGCGGCA	3240
ATATAACCAA	TAAAAACGGT	GATTTAAATA	TTACGAACGA	AGGTAGTGAT	ACTGAAATGC	3300
AAATTGGCGG	CGATGTCTCG	CAAAAAGAAG	GTAATCTCAC	GATTTCTTCT	GACAAAATCA	3360
ATATTACCAA	ACAGATAACA	ATCAAGGCAG	GTGTTGATGG	GGAGAATTCC	GATTCAGACG	3420
CGACAAACAA	TGCCAATCTA	ACCATTAAAA	CCAAAGAATT	GAAATTAACG	CAAGACCTAA	3480
ATATTTCAGG	TTTCAATAAA	GCAGAGATTA	CAGCTAAAGA	TGGTAGTGAT	TTAACTATTG	3540
GTAACACCAA	TAGTGCTGAT	GGTACTAATG	CCAAAAAAGT	AACCTTTAAC	CAGGTTAAAG	3600
ATTCAAAAAT	CTCTGCTGAC	GGTCACAAGG	TGACACTACA	CAGCAAAGTG	GAAACATCCG	3660
GTAGTAATAA	CAACACTGAA	GATAGCAGTG	ACAATAATGC	CGGCTTAACT	ATCGATGCAA	3720
AAAATGTAAC	AGTAAACAAC	AATATTACTT	CTCACAAAGC	AGTGAGCATC	TCTGCGACAA	3780
GTGGAGAAAT	TACCACTAAA	ACAGGTACAA	CCATTAACGC	AACCACTGGT	AACGTGGAGA	3840
TAACCGCTCA	AACAGGTAGT	ATCCTAGGTG	GAATTGAGTC	CAGCTCTGGC	TCTGTAACAC	3900
TTACTGCAAC	CGAGGGCGCT	CTTGCTGTAA	GCAATATTTC	GGGCAACACC	GTTACTGTTA	3960
CTGCAAATAG	CGGTGCATTA	ACCACTTTGG	CAGGCTCTAC	AATTAAAGGA	ACCGAGAGTG	4020
TAACCACTTC	AAGTCAATCA	GGCGATATCG	GCGGTACGAT	TTCTGGTGGC	ACAGTAGAGG	4080
	CGAAAGTTTA					
	AACAAGTGCA					
	AAACGCTGGC					
	AACCTTAACT					
TTACTTCAGC	CAAGGGTCAG	GTAAATCTTT	CAGCTCAGGA	TGGTAGCGTT	GCAGGAAGTA	4380

TTAATGCCGC	CAATGTGACA	CTAAATACTA	CAGGCACTTT	AACTACCGTG	AAGGGTTCAA	4440
ACATTAATGC	AACCAGCGGT	ACCTTGGTTA	TTAACGCAAA	AGACGCTGAG	CTAAATGGCG	4500
CAGCATTGGG	TAACCACACA	GTGGTAAATG	CAACCAACGC	AAATGGCTCC	GGCAGCGTAA	4560
TCGCGACAAC	CTCAAGCAGA	GTGAACATCA	CTGGGGATTT	AATCACAATA	AATGGATTAA	4620
ATATCATTTC	AAAAAACGGT	ATAAACACCG	TACTGTTAAA	AGGCGTTAAA	ATTGATGTGA	4680
AATACATTCA	ACCGGGTATA	GCAAGCGTAG	ATGAAGTAAT	TGAAGCGAAA	CGCATCCTTG	4740
agaaggtaaa	AGATTTATCT	GATGAAGAAA	GAGAAGCGTT	AGCTAAACTT	GGAGTAAGTG	4800
CTGTACGTTT	TATTGAGCCA	AATAATACAA	TTACAGTCGA	TACACAAAAT	GAATTTGCAA	4860
CCAGACCATT	AAGTCGAATA	GTGATTTCTG	AAGGCAGGGC	GTGTTTCTCA	AACAGTGATG	4920
GCGCGACGGT	GTGCGTTAAT	ATCGCTGATA	ACGGGCGGTA	GCGGTCAGTA	ATTGACAAGG	4980
TAGATTTCAT	CCTGCAATGA	AGTCATTTTA	TTTTCGTATT	ATTTACTGTG	TGGGTTAAAG	5040
TTCAGTACGG	GCTTTACCCA	TCTTGTAAAA	AATTACGGAG	AATACAATAA	AGTATTTTTA	5100
ACAGGTTATT	ATTATG					5116

#### (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1536 amino acids (B) TYPE: amino acid

  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu

Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys

Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val

Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met

Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly

G11 145	val	. Phe	e Lei	ı Ile	150	Pro	Asr	ı Gly	/ Ile	159		€ Gly	/ Lys	s Ası	160
Ile	: Ile	Asr	1 Thr	Asr 165	Gly	Phe	Thr	Ala	170		Let	ı Asp	Ile	Ser 175	
Glu	Asn	Ile	180	Ala	Arg	Asr.	Phe	Thr 185	Phe	e Glu	Glr	Thr	Lys 190		Lys
Ala	Leu	Ala 195	Glu	Ile	· Val	Asn	His 200	Gly	Leu	Ile	Thr	Val 205		Lys	Asp
Gly	Ser 210	Val	. Asn	Leu	Ile	Gly 215		Lys	Val	Lys	220		Gly	val	Il€
Ser 225	Val	Asn	Gly	Gly	Ser 230	Ile	Ser	Leu	Lev	Ala 235		Gln	Lys	Ile	Thr 240
Ile	Ser	Asp	Ile	Ile 245	Asn	Pro	Thr	Ile	Thr 250		Ser	Ile	Ala	Ala 255	
Glu	Asn	Glu	Ala 260	Val	Asn	Leu	Gly	<b>Asp</b> 265		Phe	Ala	Lys	Gly 270		Asn
Ile	Asn	Val 275	Arg	Ala	Ala	Thr	Ile 280	Arg	Asn	Gln	Gly	Lys 285	Leu	Ser	Ala
Asp	Ser 290	Val	Ser	Lys	Asp	Lys 295	Ser	Gly	Asn	Ile	Val 300		Ser	Ala	Lys
305					Ile 310					315					320
				325	Leu				330					335	
Thr	Gly	Ala	Val 340	Ile	Asp	Leu	Ser	Gly 345	Lys	Glu	Gly	Gly	Glu 350	Thr	Tyr
		355			Arg		360					365			
	370				Glu	375					380			_	-
385					Ala 390					395					400
				405	Gln				410					415	
Phe	Val	Glu	Thr 420	Ser	Gly	His	Asp	Leu 425	Phe	Ile	Lys	Asp	Asn 430	Ala	Ile
		435			Trp		440					445			
	450				Arg	455					460			_	
465					Ala 470					475					480
Thr	Leu	Thr	Asn	Thr 485	Thr	Leu	Glu	Ser	Ile 490	Leu	Lys	Lys	Gly	Thr 495	Phe

Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly 535 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu 665 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile 705 710 715 720 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Ser Val Asp Phe Thr Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn 760 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu 795 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp 810 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu 840

# SUBSTITUTE SHEET (RULE 26)

Gly	Asn 850		Thr	Ile	Asn	Asn 855		Ala	Asn	Val	Thr 860		Ile	Gly	Ser
Asp 865	Phe	Asp	Asn	His	Gln 870	Lys	Pro	Leu	Thr	Ile 875		Lys	Asp	Val	Ile 880
Ile	Asn	Ser	Gly	Asn 885	Leu	Thr	Ala	Gly	Gly 890		Ile	Val	Asn	Ile 895	
Gly	Asn	Leu	Thr 900	Val	Glu	Ser	Asn	Ala 905	Asn	Phe	Lys	Ala	Ile 910	Thr	Asn
Phe	Thr	Phe 915	Asn	Val	Gly	Gly	Leu 920	Phe	Asp	Asn	Lys	Gly 925	Asn	Ser	Asn
Ile	Ser 930	Ile	Ala	Lys	Gly	Gly 935	Ala	Arg	Phe	Lys	<b>Asp</b> 940	Ile	Asp	Asn	Ser
Lys 945	Asn	Leu	Ser	Ile	Thr 950	Thr	Asn	Ser	Ser	Ser 955	Thr	Tyr	Arg	Thr	Ile 960
Ile	Ser	Gly	Asn	Ile 965	Thr	Asn	Lys	Asn	Gly 970	Asp	Leu	Asn	Ile	Thr 975	Asn
Glu	Gly	Ser	<b>Asp</b> 980	Thr	Glu	Met	Gln	Ile 985	Gly	Gly	Asp	Val	Ser 990	Gln	Lys
Glu	Gly	Asn 995	Leu	Thr	Ile	Ser	Ser 100		Lys	Ile	Asn	Ile 100		Lys	Gln
Ile	Thr 101		Lys	Ala	Gly	Val 101		Gly	Glu	Asn	Ser 102		Ser	Asp	Ala
Thr 102		Asn	Ala	Asn	Leu 1030		Ile	Lys	Thr	Lys 103		Leu	Lys	Leu	Thr 1040
Gln	Asp	Leu	Asn	Ile 104	Ser 5	Gly	Phe	Asn	Lys 1050		Glu	Ile	Thr	Ala 105	-
Asp	Gly	Ser	Asp 1060		Thr	Ile	Gly	Asn 1065		Asn	Ser	Ala	Asp 1070		Thr
Asn	Ala	Lys 1075		Val	Thr	Phe	Asn 1080		Val	Lys	Asp	Ser 1085		Ile	Ser
Ala	Asp 1090	Gly	His	Lys	Val	Thr 1099	Leu	His	Ser	Lys	Val 1100		Thr	Ser	Gly
Ser 1109	Asn	Asn	Asn	Thr	Glu 1110		Ser	Ser	Asp	Asn 1115		Ala	Gly	Leu	Thr 1120
Ile	Asp	Ala	Lys	Asn 112	Val	Thr	Val	Asn	Asn 1130		Ile	Thr	Ser	His 1135	
Ala	Val	Ser	Ile 1140		Ala	Thr	Ser	Gly 1145		Ile	Thr	Thr	Lys 1150		Gly
Thr	Thr	Ile 1155		Ala	Thr	Thr	Gly 1160		Val	Glu	Ile	Thr 1165		Gln	Thr
Gly	Ser 1170		Leu	Gly	Gly	Ile 1175		Ser	Ser	Ser	Gly 1180		Val	Thr	Leu
Thr 1185	Ala	Thr	Glu	Gly	Ala 1190	Leu	Ala	Val	Ser	Asn 1199		Ser	Gly	Asn	Thr 1200

# SUBSTITUTE SHEET (RULE 26)

Val	Thr	Val	Thr	Ala 120		Ser	Gly	Ala	Leu 121		Thr	Leu	Ala	Gly 121	
Thr	Ile	Lys	Gly 1220		Glu	Ser	Val	Thr 122		Ser	Ser	Gln	Ser 123		Asp
Ile	Gly	Gly 1235		Ile	Ser	Gly	Gly 1240		Val	Glu	Val	Lys 124		Thr	Glu
Ser	Leu 1250	Thr	Thr	Gln	Ser	Asn 1255		Lys	Ile	Lys	Ala 1260		Thr	Gly	Glu
Ala 1265		Val	Thr	Ser	Ala 1270		Gly	Thr	Ile	Gly 127		Thr	Ile	Ser	Gly 128
Asn	Thr	Val	Asn	Val 1285		Ala	Asn	Ala	Gly 1290		Leu	Thr	Val	Gly 1295	
Gly	Ala	Glu	Ile 1300		Ala	Thr	Glu	Gly 1305		Ala	Thr	Leu	Thr 1310		Ser
Ser	Gly	Lys 1315		Thr	Thr	Glu	Ala 1320		Ser	His	Ile	Thr 1325		Ala	Lys
	Gln 1330	Val	Asn	Leu	Ser	Ala 1335		Asp	Gly	Ser	Val 1340		Gly	Ser	Ile
Asn 1345		Ala	Asn	Val	Thr 1350		Asn	Thr	Thr	Gly 1355		Leu	Thr	Thr	Val 1360
Lys	Gly	Ser	Asn	Ile 1365		Ala	Thr	Ser	Gly 1370		Leu	Val	Ile	Asn 1375	
Lys	Asp	Ala	Glu 1380		Asn	Gly	Ala	Ala 1385	Leu	Gly	Asn	His	Thr 1390		Val
Asn	Ala	Thr 1395		Ala	Asn	Gly	Ser 1400		Ser	Val	Ile	Ala 1405		Thr	Ser
Ser	Arg 1410	Val	Asn	Ile	Thr	Gly 1415		Leu	Ile	Thr	Ile 1420		Gly	Leu	Asn
Ile 1425		Ser	Lys	Asn	Gly 1430		Asn	Thr	Val	Leu 1435		Lys	Gly	Val.	Lys 144(
Ile	Asp	Val	Lys	Tyr 1445		Gln	Pro	Gly	Ile 1450		Ser	Val	Asp	Glu 1455	
Ile	Glu	Ala	Lys 1460		Ile	Leu	Glu	Lys 1465		Lys	qaA	Leu	Ser 1470	_	Glu
Glu	Arg	Glu 1475		Leu	Ala	Lys	Leu 1480		Val	Ser	Ala	Val 1485		Phe	Ile
Glu	Pro 1490	Asn	Asn	Thr	Ile	Thr 1495		Asp	Thr	Gln	Asn 1500		Phe	Ala	Thr
Arg 1505		Leu	Ser	Arg	Ile 1510		Ile	Ser	Glu	Gly 1515	_	Ala	Cys	Phe	Ser 1520
Asn	Ser	Asp	Gly	Ala 1525		Val	Cys	Val	Asn 1530		Ala	Asp	Asn	Gly 1535	

### (2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4937 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TAAATATACA	AGATAATAAA	аатааатсаа	GATTTTTGTG	ATGACAAACA	ACAATTACAA	60
CACCTTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAAAT	AGTATAAATC	CGCCATATAA	120
AATGGTATAA	TCTTTCATCT	TTCATCTTTA	ATCTTTCATC	TTTCATCTTT	CATCTTTCAT	180
CTTTCATCTT	TCATCTTTCA	TCTTTCATCT	TTCATCTTTC	ATCTTTCATC	TTTCATCTTT	240
CACATGAAAT	GATGAACCGA	GGGAAGGGAG	GGAGGGGCAA	GAATGAAGAG	GGAGCTGAAC	300
GAACGCAAAT	GATAAAGTAA	TTTAATTGTT	CAACTAACCT	TAGGAGAAAA	TATGAACAAG	360
ATATATCGTC	TCAAATTCAG	CAAACGCCTG	AATGCTTTGG	TTGCTGTGTC	TGAATTGGCA	420
CGGGGTTGTG	ACCATTCCAC	AGAAAAAGGC	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	480
CACTTAGCGT	TAAAGCCACT	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	ATCTATTCCA	540
CAATCTGTTT	TAGCAAGCGG	CTTACAAGGA	ATGGATGTAG	TACACGGCAC	AGCCACTATG	600
CAAGTAGATG	GTAATAAAAC	CATTATCCGC	AACAGTGTTG	ACGCTATCAT	TAATTGGAAA	660
CAATTTAACA	TCGACCAAAA	TGAAATGGTG	CAGTTTTTAC	AAGAAAACAA	CAACTCCGCC	720
GTATTCAACC	GTGTTACATC	TAACCAAATC	TCCCAATTAA	AAGGGATTTT	AGATTCTAAC	780
GGACAAGTCT	TTTTAATCAA	CCCAAATGGT	ATCACAATAG	GTAAAGACGC	AATTATTAAC	840
ACTAATGGCT	TTACGGCTTC	TACGCTAGAC	ATTTCTAACG	AAAACATCAA	GGCGCGTAAT	900
TTCACCTTCG	AGCAAACCAA	AGATAAAGCG	CTCGCTGAAA	TTGTGAATCA	CGGTTTAATT	960
ACTGTCGGTA	AAGACGGCAG	TGTAAATCTT	ATTGGTGGCA	aagtgaaaaa	CGAGGGTGTG	1020
ATTAGCGTAA	ATGGTGGCAG	CATTTCTTTA	CTCGCAGGGC	AAAAAATCAC	CATCAGCGAT	1080
ATAATAAACC	CAACCATTAC	TTACAGCATT	GCCGCGCCTG	AAAATGAAGC	GGTCAATCTG	1140
GGCGATATTT	TTGCCAAAGG	CGGTAACATT	AATGTCCGTG	CTGCCACTAT	TCGAAACCAA	1200
GGTAAACTTT	CTGCTGATTC	TGTAAGCAAA	GATAAAAGCG	GCAATATTGT	TCTTTCCGCC	1260
AAAGAGGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTTCCGCTC	AAAATCAGCA	AGCTAAAGGC	1320
GGCAAGCTGA	TGATTACAGG	CGATAAAGTC	ACATTAAAAA	CAGGTGCAGT	TATCGACCTT	1380
TCAGGTAAAG	AAGGGGGAGA	AACTTACCTT	GGCGGTGACG	AGCGCGGCGA	AGGTAAAAAC	1440
GGCATTCAAT	TAGCAAAGAA	AACCTCTTTA	GAAAAAGGCT	CAACCATCAA	TGTATCAGGC	1500
AAAGAAAAAG	GCGGACGCGC	TATTGTGTGG	GGCGATATTG	CGTTAATTGA	CGGCAATATT	1560
AACGCTCAAG	GTAGTGGTGA	TATCGCTAAA	ACCGGTGGTT	TTGTGGAGAC	ATCGGGGCAT	1620

TATTTATCCA	1 TTGACAGCAA	TGCAATTGTT	, <b>YYYYCYYYY</b> G	AGTGGTTGCT	AGACCCTGAT	1680
GATGTAACAA	TTGAAGCCGA	AGACCCCCTT	CGCAATAATA	CCGGTATAAA	TGATGAATTC	1740
CCAACAGGCA	CCGGTGAAGC	AAGCGACCCT	ATAAAAAAA	GCGAACTCAA	AACAACGCTA	1800
ACCAATACAA	CTATTTCAAA	TTATCTGAAA	AACGCCTGGA	CANTGAATAT	AACGGCATCA	1860
AGAAAACTTA	CCGTTAATAG	CTCAATCAAC	ATCGGAAGCA	ACTCCCACTT	AATTCTCCAT	1920
AGTAAAGGTC	AGCGTGGCGG	AGGCGTTCAG	ATTGATGGAG	ATATTACTTC	TAAAGGCGGA	1980
AATTTAACCA	TTTATTCTGG	CGGATGGGTT	GATGTTCATA	AAAATATTAC	GCTTGATCAG	2040
GGTTTTTTAA	ATATTACCGC	CGCTTCCGTA	GCTTTTGAAG	GTGGAAATAA	CAAAGCACGC	2100
GACGCGGCAA	ATGCTAAAAT	TGTCGCCCAG	GGCACTGTAA	CCATTACAGG	AGAGGGAAAA	2160
GATTTCAGGG	CTAACAACGT	ATCTTTAAAC	GGAACGGGTA	AAGGTCTGAA	TATCATTTCA	2220
TCAGTGAATA	ATTTAACCCA	CAATCTTAGT	GGCACAATTA	ACATATCTGG	GAATATAACA	2280
ATTAACCAAA	CTACGAGAAA	GAACACCTCG	TATTGGCAAA	CCAGCCATGA	TTCGCACTGG	2340
AACGTCAGTG	CTCTTAATCT	AGAGACAGGC	GCAAATTTTA	CCTTTATTAA	ATACATTTCA	2400
AGCAATAGCA	AAGGCTTAAC	AACACAGTAT	AGAAGCTCTG	CAGGGGTGAA	TTTTAACGGC	2460
GTAAATGGCA	ACATGTCATT	ÇAATCTCAAA	GAAGGAGCGA	AAGTTAATTT	CAAATTAAAA	2520
CCAAACGAGA	ACATGAACAC	AAĢCAAACCT	TTACCAATTC	GGTTTTTAGC	CAATATCACA	2580
GCCACTGGTG	GGGGCTCTGT	TTTTTTTGAT	ATATATGCCA	ACCATTCTGG	CAGAGGGGCT	2640
Gagttaaaaa	TGAGTGAAAT	TAATATCTCT	AACGGCGCTA	ATTTTACCTT	AAATTCCCAT	2700
GTTCGCGGCG	ATGACGCTTT	TAAAATCAAC	AAAGACTTAA	CCATAAATGC	AACCAATTCA	2760
AATTTCAGCC	TCAGACAGAC	GAAAGATGAT	TTTTATGACG	GGTACGCACG	CAATGCCATC	2820
AATTCAACCT	ACAACATATC	CATTCTGGGC	GGTAATGTCA	CCCTTGGTGG	ACAAAACTCA	2880
AGCAGCAGCA	TTACGGGGAA	TATTACTATC	GAGAAAGCAG	CAAATGTTAC	GCTAGAAGCC.	2940
AATAACGCCC	CTAATCAGCA	AAACATAAGG	GATAGAGTTA	TAAAACTTGG	CAGCTTGCTC	3000
GTTAATGGGA	GTTTAAGTTT	AACTGGCGAA	AATGCAGATA	TTAAAGGCAA	TCTCACTATT	3060
TCAGAAAGCG	CCACTTTTAA	AGGAAAGACT	AGAGATACCC	TAAATATCAC	CGGCAATTTT	3120
ACCAATAATG	GCACTGCCGA	AATTAATATA	ACACAAGGAG	TGGTAAAACT	TGGCAATGTT	3180
ACCAATGATG	GTGATTTAAA	CATTACCACT	CACGCTAAAC	GCAACCAAAG	AAGCATCATC	3240
GGCGGAGATA	TAATCAACAA	AAAAGGAAGC	ATTAAAATT	CAGACAGTAA	TAATGATGCT	3300
GAAATCCAAA	TTGGCGGCAA	TATCTCGCAA	AAAGAAGGCA	ACCTCACGAT	TTCTTCCGAT	3360
ATAATTAATA	TCACCAAACA	GATAACAATC	AAAAAGGGTA	TTGATGGAGA	GGACTCTAGT	3420
TCAGATGCGA	CAAGTAATGC	CAACCTAACT	ATTAAAACCA	aagaattgaa	ATTGACAGAA	3480
GACCTAAGTA	TTTCAGGTTT	CAATAAAGCA	GAGATTACAG	CCAAAGATGG	TAGAGATTTA	3540
ACTATTGGCA	ACAGTAATGA	CGGTAACAGC	GGTGCCGAAG	CCAAAACAGT	AACTTTTAAC	3600
AATGTTAAAG	ATTCAAAAAT	CTCTGCTGAC	GGTCACAATG	TGACACTAAA	TAGCAAAGTG	3660

AAAACATCTA	GCAGCAATGG	CGGACGTGAA	AGCAATAGCG	ACAACGATAC	CGGCTTAACT	372
ATTACTGCAA	AAAATGTAGA	AGTAAACAAA	GATATTACTT	CTCTCAAAAC	AGTAAATATC	378
ACCGCGTCGG	AAAAGGTTAC	CACCACAGCA	GGCTCGACCA	TTAACGCAAC	AAATGGCAAA	384
GCAAGTATTA	CAACCAAAAC	AGGTGATATC	AGCGGTACGA	TTTCCGGTAA	CACGGTAAGT	390
GTTAGCGCGA	CTGGTGATTT	AACCACTAAA	TCCGGCTCAA	AAATTGAAGC	GAAATCGGGT	396
GAGGCTAATG	TAACAAGTGC	AACAGGTACA	ATTGGCGGTA	CAATTTCCGG	TAATACGGTA	4020
AATGTTACGG	CAAACGCTGG	CGATTTAACA	GTTGGGAATG	GCGCAGAAAT	TAATGCGACA	4080
gaaggagctg	CAACCTTAAC	CGCAACAGGG	AATACCTTGA	CTACTGAAGC	CGGTTCTAGC	4140
ATCACTTCAA	CTAAGGGTCA	GGTAGACCTC	TTGGCTCAGA	ATGGTAGCAT	CGCAGGAAGC	4200
ATTAATGCTG	CTAATGTGAC	ATTAAATACT	ACAGGCACCT	TAACCACCGT	GGCAGGCTCG	4260
gatattaaag	CAACCAGCGG	CACCTTGGTT	ATTAACGCAA	AAGATGCTAA	GCTAAATGGT	4320
GATGCATCAG	GTGATAGTAC	AGAAGTGAAT	GCAGTCAACG	CAAGCGGCTC	TGGTAGTGTG	4380
ACTGCGGCAA	CCTCAAGCAG	TGTGAATATC	ACTGGGGATT	TAAACACAGT	AAATGGGTTA	4440
AATATCATTT	CGAAAGATGG	TAGAAACACT	GTGCGCTTAA	GAGGCAAGGA	AATTGAGGTG	4500
AAATATATCC	AGCCAGGTGT	AGCAAGTGTA	GAAGAAGTAA	TTGAAGCGAA	ACGCGTCCTT	4560
gaaaaagtaa	AAGATTTATC	TGATGAAGAA	AGAGAAACAT	TAGCTAAACT	TGGTGTAAGT	4620
SCTGTACGTT	TTGTTGAGCC	AAATAATACA	ATTACAGTCA	ATACACAAAA	TGAATTTACA	4680
ACCAGACCGT	CAAGTCAAGT	GATAATTTCT	GAAGGTAAGG	CGTGTTTCTC	AAGTGGTAAT	4740
GCGCACGAG	TATGTACCAA	TGTTGCTGAC	GATGGACAGC	CGTAGTCAGT	AATTGACAAG	4800
STAGATTTCA	TCCTGCAATG	AAGTCATTTT	ATTTTCGTAT	TATTTACTGT	GTGGGTTAAA	4860
STTCAGTACG	GGCTTTACCC	ATCTTGTAAA	AAATTACGGA	GAATACAATA	AAGTATTTTT	4920
ACAGGTTAT	TATTATG					4937

### (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1477 amino acids

  - (B) TYPE: amino acid
    (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu

Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys

Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly 135 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp 200 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro 250 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys 295 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp 390 395

Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys Thr Thr Leu Thr Asn Thr Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp 485 490 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile 505 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg 520 Gly Gly Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Gly Asn 535 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr 555 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu 570 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala 585 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly 630 Asn Ile Thr Ile Asn Gln Thr Thr Arg Lys Asn Thr Ser Tyr Trp Gln Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr 665 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly 680 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Gly Ala Lys Val Asn Phe Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Ser Val Phe Phe 740 745

Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala 790 795 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ser Ile Thr 840 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg 950 955 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr 1000 1005 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys 1030 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr 1045 1050 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn 1065 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys 1090 1095 1100

Thr 110	Ser	Ser	Ser	Asn	Gly 111	Gly 0	Arg	Glu	Ser	Asn 111		Asp	Asn	Asp	Thr 1120
Gly	Leu	Thr	Ile	Thr 112	Ala 5	Lys	Asn	Val	Glu 113		Asn	Lys	Asp	Ile 113	
Ser	Leu	Lys	Thr 114	Val 0	Asn	Ile	Thr	Ala 114		Glu	Lys	Val	Thr 115		Thr
Ala	Gly	Ser 115	Thr 5	Ile	Asn	Ala	Thr 116		Gly	Lys	Ala	Ser 116		Thr	Thr
Lys	Thr 117	Gly 0	Asp	Ile	Ser	Gly 117		Ile	Ser	Gly	Asn 118		Val	Ser	Val
Ser 118	Ala 5	Thr	Val	Asp	Leu 119	Thr 0	Thr	Lys	Ser	Gly 119		Lys	Ile	Glu	Ala 1200
Lys	Ser	Gly	Glu	Ala 120	Asn 5	Val	Thr	Ser	Ala 121		Gly	Thr	Ile	Gly 121	
Thr	Ile	Ser	Gly 122		Thr	Val	Asn	Val 122		Ala	Asn	Ala	Gly 123		Leu
Thr	Val	Gly 123	Asn 5	Gly	Ala	Glu	Ile 124		Ala	Thr	Glu	Gly 124		Ala	Thr
Leu	Thr 125	Ala	Thr	Gly	Asn	Thr 125	Leu 5	Thr	Thr	Glu	Ala 1260		Ser	Ser	Ile
Thr 126	Ser 5	Thr	Lys	Gly	Gln 127		Asp	Leu	Leu	Ala 1275		Asn	Gly	Ser	Ile 1280
Ala	Gly	Ser	Ile	Asn 1285		Ala	Asn	Val	Thr 1290		Asn	Thr	Thr	Gly 1299	
Leu	Thr	Thr	Val 1300	Ala O	Gly	Ser	Asp	Ile 1305		Ala	Thr	Ser	Gly 1310		Leu
Val	Ile	Asn 1315	Ala	Lys	Asp	Ala	Lys 1320		Asn	Gly	qaA	Ala 1325		Gly	Asp
Ser	Thr 1330	Glu	Val	Asn	Ala	Val 1335	Asn	Ala	Ser	Gly	Ser 1340		Ser	Val	Thr
Ala 1345	Ala	Thr	Ser	Ser	Ser 1350	Val	Asn	Ile	Thr	Gly 1355			Asn		Val 1360
Asn	Gly	Leu	Asn	Ile 1365	Ile	Ser	Lys	Asp	Gly 1370		Asn	Thr	Val	Arg 1375	
Arg	Gly	Lys	Glu 1380	Ile	Glu	Val	Lys	Tyr 1385		Gln	Pro	Gly	Val 1390		Ser
Val	Glu	Glu 1395	Val	Ile	Glu	Ala	Lys 1400		Val	Leu	Glu	Lys 1405		Lys	Asp
Leu	Ser 1410	Asp	Glu	Glu	Arg	Glu 1415	Thr	Leu	Ala	Lys	Leu 1420		Val	Ser	Ala
Val 1425	Arg	Phe	Val	Glu	Pro 1430		Asn	Thr	Ile	Thr 1435		Asn	Thr	Gln	Asn 1440
Glu	Phe	Thr	Thr	Arg 1445		Ser	Ser	Gln	Val 1450		Ile	Ser	Glu	Gly 1455	

Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala 1465

Asp Asp Gly Gln Pro 1475

### (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9171 base pairs

  - (B) TYPE: nucleic acid (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAGCGTTCT	CTTAATACTA	GTACAAACCC	ACAATAAAAT	ATGACAAACA	ACAATTACAA	60
CACCTTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAATA	GTATAAATCC	GCCATATAAA	120
ATGGTATAAT	CTTTCATCTT	TCATCTTTCA	TCTTTCATCT	TTCATCTTTC	ATCTTTCATC	180
TTTCATCTTT	CATCTTTCAT	CTTTCATCTT	TCATCTTTCA	TCTTTCATCT	TTCATCTTTC	240
ACATGAAATG	ATGAACCGAG	GGAAGGGAGG	GAGGGGCAAG	AATGAAGAGG	GAGCTGAACG	300
AACGCAAATG	ATAAAGTAAT	TTAATTGTTC	AACTAACCTT	AGGAGAAAAT	ATGAACAAGA	360
TATATCGTCT	CAAATTCAGC	AAACGCCTGA	ATGCTTTGGT	TGCTGTGTCT	GAATTGGCAC	420
GGGGTTGTGA	CCATTCCACA	GAAAAAGGCA	GCGAAAAACC	TGCTCGCATG	AAAGTGCGTC	480
ACTTAGCGTT	AAAGCCACTT	TCCGCTATGT	TACTATCTTT	AGGTGTAACA	TCTATTCCAC	540
AATCTGTTTT	AGCAAGCGGC	TTACAAGGAA	TGGATGTAGT	ACACGGCACA	GCCACTATGC	600
AAGTAGATGG	TAATAAAACC	ATTATCCGCA	ACAGTGTTGA	CGCTATCATT	AATTGGAAAC	660
AATTTAACAT	CGACCAAAAT	GAAATGGTGC	AGTTTTTACA	AGAAAACAAC	AACTCCGCCG	720
TATTCAACCG	TGTTACATCT	AACCAAATCT	CCCAATTAAA	AGGGATTTTA	GATTCTAACG	780
GACAAGTCTT	TTTAATCAAC	CCAAATGGTA	TCACAATAGG	TAAAGACGCA	ATTATTAACA	840
CTAATGGCTT	TACGGCTTCT	ACGCTAGACA	TTTCTAACGA	AAACATCAAG	GCGCGTAATT	900
TCACCTTCGA	GCAAACCAAA	GATAAAGCGC	TCGCTGAAAT	TGTGAATCAC	GGTTTAATTA	960
CTGTCGGTAA	AGACGGCAGT	GTAAATCTTA	TTGGTGGCAA	AGTGAAAAAC	GAGGGTGTGA	1020
TTAGCGTAAA	TGGTGGCAGC	ATTTCTTTAC	TCGCAGGGCA	AAAAATCACC	ATCAGCGATA	1080
TAATAAACCC	AACCATTACT	TACAGCATTG	CCGCGCCTGA	AAATGAAGCG	GTCAATCTGG	1140
GCGATATTTT	TGCCAAAGGC	GGTAACATTA	ATGTCCGTGC	TGCCACTATT	CGAAACCAAG	1200
CTTTCCGCCA	aagagggtga	AGCGGAAATT	GGCGGTGTAA	TTTCCGCTCA	AAATCAGCAA	1260
GCTAAAGGCG	GCAAGCTGAT	GATTACAGGC	GATAAAGTCA	CATTAAAAAC	AGGTGCAGTT	1320
ATCGACCTTT	CAGGTAAAGA	AGGGGGAGAA	ACTTACCTTG	GCGGTGACGA	GCGCGGCGAA	1380
GGTAAAAACG	GCATTCAATT	AGCAAAGAAA	ACCTCTTTAG	AAAAAGGCTC	AACCATCAAT	1440

GTATCAGGCA	A AAGAAAAAGG	CGGACGCGCT	ATTGTGTGGG	GCGATATTGC	GTTAATTGAC	1500
GGCAATATTA	ACGCTCAAGG	TAGTGGTGAT	ATCGCTAAAA	CCGGTGGTTT	TGTGGAGACG	1560
TCGGGGCATG	ATTTATTCAT	CAAAGACAAT	GCAATTGTTG	ACGCCAAAGA	GTGGTTGTTA	1620
GACCCGGATA	ATGTATCTAT	TAATECAGAA	ACAGCAGGAC	GCAGCAATAC	TTCAGAAGAC	1680
GATGAATACA	CGGGATCCGG	GAATAGTGCC	AGCACCCCAA	AACGAAACAA	AGAAAAGACA	1740
ACATTAACAA	ACACAACTCT	TGAGAGTATA	CTAAAAAAAG	GTACCTTTGT	TAACATCACT	1800
GCTAATCAAC	GCATCTATGT	CAATAGCTCC	ATTAATTTAT	CCAATGGCAG	CTTAACTCTT	1860
TGGAGTGAGG	GTCGGAGCGG	TGGCGGCGTT	GAGATTAACA	ACGATATTAC	CACCGGTGAT	1920
GATACCAGAG	GTGCAAACTT	AACAATTTAC	TCAGGCGGCT	GGGTTGATGT	TCATAAAAAT	1980
ATCTCACTCG	GGGCGCAAGG	TAACATAAAC	ATTACAGCTA	AACAAGATAT	CGCCTTTGAG	2040
aaaggaagca	ACCAAGTCAT	TACAGGTCAA	GGGACTATTA	CCTCAGGCAA	TCAAAAAGGT	2100
TTTAGATTTA	ATAATGTCTC	TCTAAACGGC	ACTGGCAGCG	GACTGCAATT	CACCACTAAA	2160
AGAACCAATA	AATACGCTAT	CACAAATAAA	TTTGAAGGGA	CTTTAAATAT	TTCAGGGAAA	2220
GTGAACATCT	CAATGGTTTT	ACCTAAAAAT	GAAAGTGGAT	ATGATAAATT	CAAAGGACGC	2280
acttactgga	ATTTAACCTC	GAAAGTGGAT	ATGATAAATT	CAAAGGACGC	CCTCACTATT	2340
GACTCCAGAG	GAAGCGATAG	TGCAGGCACA	CTTACCCAGC	CTTATAATTT	AAACGGTATA	2400
TCATTCAACA	AAGACACTAC	CTTTAATGTT	GAACGAAATG	CAAGAGTCAA	CTTTGACATC	2460
AAGGCACCAA	TAGGGATAAA	TAAGTATTCT	AGTTTGAATT	ACGCATCATT	TAATGGAAAC	2520
ATTTCAGTTT	CGGGAGGGG	GAGTGTTGAT	TTCACACTTC	TCGCCTCATC	CTCTAACGTC	2580
CAAACCCCCG	GTGTAGTTAT	AAATTCTAAA	TACTTTAATG	TTTCAACAGG	GTCAAGTTTA	2640
AGATTTAAAA	CTTCAGGCTC	AACAAAAACT	GGCTTCTCAA	TAGAGAAAGA	TTTAACTTTA	2700
AATGCCACCG	GAGGCAACAT	AACACTTTTG	CAAGTTGAAG	GCACCGATGG	AATGATTGGT	2760
AAAGGCATTG	TAGCCAAAAA	AAACATAACC	TTTGAAGGAG	GTAAGATGAG	GTTTGGCTCC	2820
AGGAAAGCCG	TAACAGAAAT	CGAAGGCAAT	GTTACTATCA	ATAACAACGC	TAACGTCACT	2880
CTTATCGGTT	CGGATTTTGA	CAACCATCAA	AAACCTTTAA	CTATTAAAAA	AGATGTCATC	2940
ATTAATAGCG	GCAACCTTAC	CGCTGGAGGC	AATATTGTCA	ATATAGCCGG	AAATCTTACC	3000
GTTGAAAGTA	ACGCTAATTT	CAAAGCTATC	ACAAATTTCA	CTTTTAATGT	AGGCGGCTTG	3060
TTTGACAACA	AAGGCAATTC	AAATATTTCC	ATTGCCAAAG	GAGGGGCTCG	CTTTAAAGAC	3120
ATTGATAATT	CCAAGAATTT	AAGCATCACC	ACCAACTCCA	GCTCCACTTA	CCGCACTATT	3180
ATAAGCGGCA	ATATAACCAA	TAAAAACGGT	GATTTAAATA	TTACGAACGA	AGGTAGTGAT	3240
ACTGAAATGC	AAATTGGCGG	CGATGTCTCG	CAAAAAGAAG	GTAATCTCAC	GATTTCTTCT	3300
GACAAAATCA	ATATTACCAA	ACAGATAACA	ATCAAGGCAG	GTGTTGATGG	GGAGAATTCC	3360
GATTCAGACG	CGACAAACAA	TGCCAATCTA	ACCATTAAAA	CCAAAGAATT	GAAATTAACG	3420
CAAGACCTAA	ATATTTCAGG	TTTCAATAAA	GCAGAGATTA	CAGCTAAAGA	TGGTAGTGAT	3480

TTAACTATTG	GTAACACCAA	TAGTGCTGAT	GGTACTAATG	CCAAAAAAGT	AACCTTTAAC	3540
CAGGTTAAAG	ATTCAAAAAT	CTCTGCTGAC	GGTCACAAGG	TGACACTACA	CAGCAAAGTG	3600
GANACATCCG	GTAGTAATAA	CAACACTGAA	GATAGCAGTG	ACAATAATGC	CGGCTTAACT	3660
ATCGATGCAA	AAAATGTAAC	AGTAAACAAC	AATATTACTT	CTCACAAAGC	AGTGAGCATC	3720
TCTGCGACAA	GTGGAGAAAT	TACCACTAAA	ACAGGTACAA	CCATTAACGC	AACCACTGGT	3780
AACGTGGAGA	TAACCGCTCA	AACAGGTAGT	ATCCTAGGTG	GAATTGAGTC	CAGCTCTGGC	3840
TCTGTAACAC	TTACTGCAAC	CGAGGGCGCT	CTTGCTGTAA	GCAATATTTC	GGGCAACACC	3900
GTTACTGTTA	CTGCAAATAG	CGGTGCATTA	ACCACTTTGG	CAGGCTCTAC	AATTAAAGGA	3960
ACCGAGAGTG	TAACCACTTC	AAGTCAATCA	GGCGATATCG	GCGGTACGAT	TTCTGGTGGC	4020
ACAGTAGAGG	TTAAAGCAAC	CGAAAGTTTA	ACCACTCAAT	CCAATTCAAA	AATTAAAGCA	4080
ACAACAGGCG	AGGCTAACGT	AACAAGTGCA	ACAGGTACAA	TTGGTGGTAC	GATTTCCGGT	4140
AATACGGTAA	ATGTTACGGC	AAACGCTGGC	GATTTAACAG	TTGGGAATGG	CGCAGAAATT	4200
AATGCGACAG	AAGGAGCTGC	AACCTTAACT	ACATCATCGG	GCAAATTAAC	TACCGAAGCT	4260
AGTTCACACA	TTACTTCAGC	CAAGGGTCAG	GTAAATCTTT	CAGCTCAGGA	TGGTAGCGTT	4320
GCAGGAAGTA	TTAATGCCGC	CAATGTGACA	CTAAATACTA	CAGGCACTTT	AACTACCGTG	4380
AAGGGTTCAA	ACATTAATGC	AACCAGCGGT	ACCTTGGTTA	TTAACGCAAA	AGACGCTGAG	4440
CTAAATGGCG	CAGCATTGGG	TAACCACACA	GTGGTAAATG	CAACCAACGC	AAATGGCTCC	4500
GGCAGCGTAA	TCGCGACAAC	CTCAAGCAGA	GTGAACATCA	CTGGGGATTT	AATCACAATA	4560
AATGGATTAA	ATATCATTTC	AAAAAACGGT	ATAAACACCG	TACTGTTAAA	aggcgttaaa	4620
ATTGATGTGA	AATACATTCA	ACCGGGTATA	GCAAGCGTAG	ATGAAGTAAT	TGAAGCGAAA	4680
CGCATCCTTG	AGAAGGTAAA	AGATTTATCT	GATGAAGAAA	GAGAAGCGTT	AGCTAAACTT	4740
GGCGTAAGTG	CTGTACGTTT	TATTGAGCCA	AATAATACAA	TTACAGTCGA	TACACAAAAT	4800
GAATTTGCAA	CCAGACCATT	AAGTCGAATA	GTGATTTCTG	AAGGCAGGGC	GTGTTTCTCA	4860
AACAGTGATG	GCGCGACGGT	GTGCGTTAAT	ATCGCTGATA	ACGGGCGGTA	GCGGTCAGTA	4920
ATTGACAAGG	TAGATTTCAT	CCTGCAATGA	AGTCATTTTA	TTTTCGTATT	ATTTACTGTG	4980
TGGGTTAAAG	TTCAGTACGG	GCTTTACCCA	TCTTGTAAAA	AATTACGGAG	AATACAATAA	5040
AGTATTTTTA	ACAGGTTATT	ATTATGAAAA	ATATAAAAAG	CAGATTAAAA	CTCAGTGCAA	5100
TATCAGTATT	GCTTGGCCTG	GCTTCTTCAT	CATTGTATGC	AGAAGAAGCG	TTTTTAGTAA	5160
AAGGCTTTCA	GTTATCTGGT	GCACTTGAAA	CTTTAAGTGA	AGACGCCCAA	CTGTCTGTAG	5220
CAAAATCTTT	ATCTAAATAC	CAAGGCTCGC	AAACTTTAAC	AAACCTAAAA	ACAGCACAGC	5280
TTGAATTACA	GGCTGTGCTA	GATAAGATTG	AGCCAAATAA	GTTTGATGTG	ATATTGCCAC	5340
AACAAACCAT	TACGGATGGC	AATATTATGT	TTGAGCTAGT	CTCGAAATCA	GCCGCAGAAA	5400
GCCAAGTTTT	TTATAAGGCG	AGCCAGGGTT	ATAGTGAAGA	AAATATCGCT	CGTAGCCTGC	5460
CATCTTTGAA	ACAAGGAAAA	GTGTATGAAG	ATGGTCGTCA	GTGGTTCGAT	TTGCGTGAAT	5520

TCAATATGGC	AAAAGAAAAT	CCACTTAAAG	TCACTCGCGT	GCATTACGAG	TTAAACCCTA	5580
AAAACAAAAC	CTCTGATTTG	GTAGTTGCAG	GTTTTTCGCC	TTTTGGCAAA	ACGCGTAGCT	5640
TTGTTTCCTA	TGATAATTTC	GGCGCAAGGG	AGTTTAACTA	TCAACGTGTA	AGTCTAGGTT	5700
TTGTAAATGC	CAATTTGACC	GGACATGATG	ATGTATTAAA	TCTAAACCCA	TTGACCAATG	5760
TAAAAGCACC	ATCAAAATCT	TATGCGGTAG	GCATAGGATA	TACTTATCCG	TTTTATGATA	5820
AACACCAATC	CTTAAGTCTT	TATACCAGCA	TGAGTTATGC	TGATTCTAAT	GATATCGACG	5880
GCTTACCAAG	TGCGATTAAT	CGTAAATTAT	CAAAAGGTCA	ATCTATCTCT	GCGAATCTGA	5940
aatggagtta	TTATCTCCCG	ACATTTAACC	TTGGAATGGA	AGACCAGTTT	AAAATTAATT	6000
TAGGCTACAA	CTACCGCCAT	ATTAATCAAA	CATCCGAGTT	AAACACCCTG	GGTGCAACGA	6060
AGAAAAATT	TGCAGTATCA	GGCGTAAGTG	CAGGCATTGA	TGGACATATC	CAATTTACCC	6120
CTAAAACAAT	CTTTAATATT	GATTTAACTC	ATCATTATTA	CGCGAGTAAA	TTACCAGGCT	6180
CTTTTGGAAT	GGAGCGCATT	GGCGAAACAT	TTAATCGCAG	CTATCACATT	AGCACAGCCA	6240
GTTTAGGGTT	GAGTCAAGAG	TTTGCTCAAG	GTTGGCATTT	TAGCAGTCAA	TTATCGGGTC	6300
AGTTTACTCT	ACAAGATATA	AGTAGCATAG	ATTTATTCTC	TGTAACAGGT	ACTTATGGCG	6360
TCAGAGGCTT	TAAATACGGC	GGTGCAAGTG	GTGAGCGCGG	TCTTGTATGG	CGTAATGAAT	6420
TAAGTATGCC	AAAATACACC	CGCTTTCAAA	TCAGCCCTTA	TGCGTTTTAT	GATGCAGGTC	6480
AGTTCCGTTA	TAATAGCGAA	AATGCTAAAA	CTTACGGCGA	AGATATGCAC	ACGGTATCCT	6540
CTGCGGGTTT	AGGCATTAAA	ACCTCTCCTA	CACAAAACTT	AAGCTTAGAT	GCTTTTGTTG	6600
CTCGTCGCTT	TGCAAATGCC	AATAGTGACA	ATTTGAATGG	CAACAAAAA	CGCACAAGCT	6660
CACCTACAAC	CTTCTGGGGT	AGATTAACAT	TCAGTTTCTA	ACCCTGAAAT	TTAATCAACT	6720
GGTAAGCGTT	CCGCCTACCA	GTTTATAACT	ATATGCTTTA	CCCGCCAATT	TACAGTCTAT	6780
ACGCAACCCT	GTTTTCATCC	TTATATATCA	AACAAACTAA	GCAAACCAAG	CAAACCAAGC	6840
AAACCAAGCA	AACCAAGCAA	ACCAAGCAAA	CCAAGCAAAC	CAAGCAAACC	AAGCAAACCA	6900
AGCAAACCAA	GCAAACCAAG	CAAACCAAGC	AAACCAAGCA	ATGCTAAAAA	ACAATTTATA	6960
TGATAAACTA	AAACATACTC	CATACCATGG	CAATACAAGG	GATTTAATAA	TATGACAAAA	7020
GAAAATTTAC	AAAGTGTTCC	ACAAAATACG	ACCGCTTCAC	TTGTAGAATC	AAACAACGAC	7080
CAAACTTCCC	TGCAAATACT	TAAACAACCA	CCCAAACCCA	ACCTATTACG	CCTGGAACAA	7140
CATGTCGCCA	AAAAAGATTA	TGAGCTTGCT	TGCCGCGAAT	TAATGGCGAT	TTTGGAAAAA	7200
ATGGACGCTA	ATTTTGGAGG	CGTTCACGAT	ATTGAATTTG	ACGCACCTGC	TCAGCTGGCA	7260
TATCTACCCG	AAAAACTACT	AATTCATTTT	GCCACTCGTC	TCGCTAATGC	AATTACAACA	7320
CTCTTTTCCG	ACCCCGAATT	GGCAATTTCC	GAAGAAGGGG	CATTAAAGAT	GATTAGCCTG	7380
CAACGCTGGT	TGACGCTGAT	TTTTGCCTCT	TCCCCCTACG	TTAACGCAGA	CCATATTCTC	7440
ATAAATATA	ATATCAACCC	AGATTCCGAA	GGTGGCTTTC	ATTTAGCAAC	AGACAACTCT	7500
דירידיא ישיריבירייה	እ እ <b>ጥጥርጥ</b> ርጥ እ ጥ	THE PROPERTY OF THE PARTY OF TH	CCCCNATCCN	አጥርጥር እአጥአጥ	CACTUTACAT	7560

GCGTTATGGG	CAGGGAATCA	ACAACTTTGT	GCTTCATTGT	GTTTTGCGTT	GCAGTCTTCA	7620
CGTTTTATTC	GTACTGCATC	TGCGTTTCAT	AAAAGAGCGG	TGGTTTTACA	GTGGTTTCCT	7680
AAAAAACTCG	CCGAAATTGC	TAATTTAGAT	GAATTGCCTG	CAAATATCCT	TCATGATGTA	7740
TATATGCACT	GCAGTTATGA	TTTAGCAAAA	AACAAGCACG	ATGTTAAGCG	TCCATTAAAC	7800
GAACTTGTCC	GCAAGCATAT	CCTCACGCAA	GGATGGCAAG	ACCGCTACCT	TTACACCTTA	7860
GGTAAAAAGG	ACGGCAAACC	TGTGATGATG	GTACTGCTTG	AACATTTTAA	TTCGGGACAT	7920
TCGATTTATC	GCACGCATTC	AACTTCAATG	ATTGCTGCTC	GAGAAAAATT	CTATTTAGTC	7980
GGCTTAGGCC	ATGAGGGCGT	TGATAACATA	GGTCGAGAAG	TGTTTGACGA	GTTCTTTGAA	8040
ATCAGTAGCA	ATAATATAAT	GGAGAGACTG	TTTTTTATCC	GTAAACAGTG	CGAAACTTTC	8100
CAACCCGCAG	TGTTCTATAT	GCCAAGCATT	GGCATGGATA	TTACCACGAT	TTTTGTGAGC	8160
AACACTCGGC	TTGCCCCTAT	TCAAGCTGTA	GCCTTGGGTC	ATCCTGCCAC	TACGCATTCT	8220
GAATTTATTG	ATTATGTCAT	CGTAGAAGAT	GATTATGTGG	GCAGTGAAGA	TTGTTTTAGC	8280
GAAACCCTTT	TACGCTTACC	CAAAGATGCC	CTACCTTATG	TACCATCTGC	ACTCGCCCCA	8340
Caaaaagtgg	ATTATGTACT	CAGGGAAAAC	CCTGAAGTAG	TCAATATCGG	TATTGCCGCT	8400
ACCACAATGA	AATTAAACCC	TGAATTTTTG	CTAACATTGC	AAGAAATCAG	AGATAAAGCT	8460
aaagtcaaaa	TACATTTTCA	TTTCGCACTT	GGACAATCAA	CAGGCTTGAC	ACACCCTTAT	8520
etcaaatggt	TTATCGAAAG	CTATTTAGGT	GACGATGCCA	CTGCACATCC	CCACGCACCT	8580
<b>FATCACGATT</b>	ATCTGGCAAT	ATTGCGTGAT	TGCGATATGC	TACTAAATCC	GTTTCCTTTC	8640
GTAATACTA	ACGGCATAAT	TGATATGGTT	ACATTAGGTT	TAGTTGGTGT	ATGCAAAACG	8700
eggga <b>tgaa</b> g	TACATGAACA	TATTGATGAA	GGTCTGTTTA	AACGCTTAGG	ACTACCAGAA	8760
rggctgatag	CCGACACACG	AGAAACATAT	ATTGAATGTG	CTTTGCGTCT	AGCAGAAAAC	8820
CATCAAGAAC	GCCTTGAACT	CCGTCGTTAC	ATCATAGAAA	ACAACGGCTT	ACAAAAGCTT	8880
TTACAGGCG	ACCCTCGTCC	ATTGGGCAAA	ATACTGCTTA	AGAAAACAAA	TGAATGGAAG	8940
EGGAAGCACT	TGAGTAAAAA	ATAACGGTTT	TTTAAAGTAA	AAGTGCGGTT	AATTTTCAAA	9000
CGTTTTAAA	AACCTCTCAA	AAATCAACCG	CACTTTTATC	TTTATAACGC	TCCCGCGCGC	9060
TGACAGTTTA	TCTCTTTCTT	AAAATACCCA	TAAAATTGTG	GCAATAGTTG	GGTAATCAAA	9120
TCAATTGTT	GATACGGCAA	ACTAAAGACG	GCGCGTTCTT	CGGCAGTCAT	С	9171

### (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9323 base pairs

  - (B) TYPE: nucleic acid (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi)	SEOUENCE	DESCRIPTION:	SEO	ID	NO:6:

CGCCACTTCA	ATTTTGGATT	GTTGAAATTC	AACTAACCAA	AAAGTGCGGT	TAAAATCTGT	60
GGAGAAAATA	GGTTGTAGTG	AAGAACGAGG	TAATTGTTCA	AAAGGATAAA	GCTCTCTTAA	120
TTGGGCATTG	GTTGGCGTTT	CTTTTTCGGT	TAATAGTAAA	TTATATTCTG	GACGACTATG	180
CAATCCACCA	ACAACTTTAC	CGTTGGTTTT	AAGCGTTAAT	GTAAGTTCTT	GCTCTTCTTG	240
GCGAATACGT	AATCCCATTT	TTTGTTTAGC	AAGAAAATGA	TCGGGATAAT	CATAATAGGT	300
GTTGCCCAAA	AATAAATTT	GATGTTCTAA	AATCATAAAT	TTTGCAAGAT	ATTGTGGCAA	360
TTCAATACCT	ATTTGTGGCG	AAATCGCCAA	TTTTAATTCA	ATTTCTTGTA	GCATAATATT	420
TCCCACTCAA	ATCAACTGGT	TAAATATACA	AGATAATAAA	AATAAATCAA	GATTTTTGTG	480
ATGACAAACA	ACAATTACAA	CACCTTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAAAT	540
AGTATAAATC	CGCCATATAA	AATGGTATAA	TCTTTCATCT	TTCATCTTTC	ATCTTTCATC	600
TTTCATCTTT	CATCTTTCAT	CTTTCATCTT	TCATCTTTCA	TCTTTCATCT	TTCATCTTTC	660
ATCTTTCATC	TTTCATCTTT	CACATGAAAT	GATGAACCGA	GGGAAGGGAG	GGAGGGGCAA	720
GAATGAAGAG	GGAGCTGAAC	GAACGCAAAT	GATAAAGTAA	TTTAATTGTT	CAACTAACCT	780
TAGGAGAAAA	TATGAACAAG	ATATATCGTC	TCAAATTCAG	CAAACGCCTG	AATGCTTTGG	840
TTGCTGTGTC	TGAATTGGCA	CGGGGTTGTG	ACCATTCCAC	AGAAAAAGGC	AGCGAAAAAC	900
CTGCTCGCAT	GAAAGTGCGT	CACTTAGCGT	TAAAGCCACT	TTCCGCTATG	TTACTATCTT	960
TAGGTGTAAC	ATCTATTCCA	CAATCTGTTT	TAGCAAGCGG	CAATTTAACA	TCGACCAAAA	1020
TGAAATGGTG	CAGTTTTTAC	AAGAAAACAA	GTAATAAAAC	CATTATCCGC	AACAGTGTTG	1080
ACGCTATCAT	TAATTGGAAA	CAATTTAACA	TCGACCAAAA	TGAAATGGTG	CAGTTTTTAC	1140
AAGAAAACAA	CAACTCCGCC	GTATTCAACC	GTGTTACATC	TAACCAAATC	TCCCAATTAA	1200
AAGGGATTTT	AGATTCTAAC	GGACAAGTCT	TTTTAATCAA	CCCAAATGGT	ATCACAATAG	1260
GTAAAGACGC	AATTATTAAC	ACTAATGGCT	TTACGGCTTC	TACGCTAGAC	ATTTCTAACG	1320
AAAACATCAA	GGCGCGTAAT	TTCACCTTCG	AGCAAACCAA	AGATAAAGCG	CTCGCTGAAA	1380
TTGTGAATCA	CGGTTTAATT	ACTGTCGGTA	AAGACGGCAG	TGTAAATCTT	ATTGGTGGCA	1440
aagtgaaaaa	CGAGGGTGTG	ATTAGCGTAA	ATGGTGGCAG	CATTTCTTTA	CTCGCAGGGC	1500
AAAAAATCAC	CATCAGCGAT	ATAATAAACC	CAACCATTAC	TTACAGCATT	GCCGCGCCTG	1560
AAAATGAAGC	GGTCAATCTG	GGCGATATTT	TTGCCAAAGG	CGGTAACATT	AATGTCCGTG	1620
CTGCCACTAT	TCGAAACCAA	GGTAAACTTT	CTGCTGATTC	TGTAAGCAAA	GATAAAAGCG	1680
GCAATATTGT	TCTTTCCGCC	AAAGAGGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTTCCGCTC	1740
AAAATCAGCA	AGCTAAAGGC	GGCAAGCTGA	TGATAAAGTC	CGATAAAGTC	ACATTAAAAA	1800
CAGGTGCAGT	TATCGACCTT	TCAGGTAAAG	AAGGGGGAGA	AACTTACCTT	GGCGGTGACG	1860
AGCGCGGCGA	AGGTAAAAAC	GGCATTCAAT	TAGCAAAGAA	AACCTCTTTA	Gaaaaaggct	1920
CAACCATCAA	TGTATCAGGC	AAAGAAAAG	GCGGACGCGC	TATTGTGTGG	GGCGATATTG	1980

CGTTAATTGA	CGGCAATATT	AACGCTCAAG	GTAGTGGTGA	TATCGCTAAA	ACCGGTGGTT	2040
TTGTGGAGAC	ATCGGGGCAT	TATTTATCCA	TTGACAGCAA	TGCAATTGTT	AAAACAAAAG	2100
AGTGGTTGCT	AGACCCTGAT	GATGTAACAA	TTGAAGCCGA	AGACCCCCTT	CGCAATAATA	2160
CCGGTATAAA	TGATGAATTC	CCAACAGGCA	CCGGTGAAGC	AAGCGACCCT	Алааааата	2220
GCGAACTCAA	AACAACGCTA	ACCAATACAA	CTATTTCAAA	TTATCTGAAA	AACGCCTGGA	2280
CAATGAATAT	AACGGCATCA	AGAAAACTTA	CCGTTAATAG	CTCAATCAAC	ATCGGAAGCA	2340
ACTCCCACTT	AATTCTCCAT	AGTAAAGGTC	AGCGTGGCGG	AGGCGTTCAG	ATTGATGGAG	2400
ATATTACTTC	TAAAGGCGGA	AATTTAACCA	TTTATTCTGG	CGGATGGGTT	GATGTTCATA	2460
AAAATATTAC	GCTTGATCAG	GGTTTTTTAA	ATATTACCGC	CGCTTCCGTA	GCTTTTGAAG	2520
GTGGAAATAA	CAAAGCACGC	GACGCGGCAA	ATGCTAAAAT	TGTCGCCCAG	GGCACTGTAA	2580
CCATTACAGG	AGAGGGAAAA	GATTTCAGGG	CTAACAACGT	ATCTTTAAAC	GGAACGGGTA	2640
AAGGTCTGAA	TATCATTTCA	TCAGTGAATA	ATTTAACCCA	CAATCTTAGT	GGCACAATTA	2700
ACATATCTGG	GAATATAACA	ATTAACCAAA	CTACGAGAAA	GAACACCTCG	TATTGGCAAA	2760
CCAGCCATGA	TTCGCACTGG	AACGTCAGTG	CTCTTAATCT	AGAGACAGGC	GCAAATTTTA	2820
CCTTTATTAA	ATACATTTCA	AGCAATAGCA	AAGGCTTAAC	AACACAGTAT	AGAAGCTCTG	2880
CAGGGGTGAA	TTTTAACGGC	GTAAATGGCA	ACATGTCATT	CAATCTCAAA	GAAGGAGCGA	2940
AAGTTAATTT	CAAATTAAAA	CCAAACGAGA	ACATGAACAC	AAGCAAACCT	TTACCAATTC	3000
GGTTTTTAGC	CAATATCACA	GCCACTGGTG	GGGGCTCTGT	TTTTTTTGAT	ATATATGCCA	3060
ACCATTCTGG	CAGAGGGGCT	GAGTTAAAAA	TGAGTGAAAT	TAATATCTCT	AACGGCGCTA	3120
ATTTTACCTT	AAATTCCCAT	GTTCGCGGCG	ATGACGCTTT	TAAAATCAAC	AAAGACTTAA	3180
CCATAAATGC	AACCAATTCA	AATTTCAGCC	TCAGACAGAC	GAAAGATGAT	TTTTATGACG	3240
GGTACGCACG	CAATGCCATC	AATTCAACCT	ACAACATATC	CATTCTGGGC	GGTAATGTCA	3300
CCCTTGGTGG	ACAAAACTCA	AGCAGCAGCA	TTACGGGGAA	TATTACTATC	GAGAAAGCAG	3360
CAAATGTTAC	GCTAGAAGCC	AATAACGCCC	CTAATCAGCA	AAACATAAGG	GATAGAGTTA	3420
TAAAACTTGG	CAGCTTGCTC	GTTAATGGGA	GTTTAAGTTT	AACTGGCGAA	AATGCAGATA	3480
TTAAAGGCAA	TCTCACTATT	TCAGAAAGCG	CCACTTTTAA	AGGAAAGACT	AGAGATACCC	3540
TAAATATCAC	CGGCAATTTT	ACCAATAATG	GCACTGCCGA	ATTAATATA	ACACAAGGAG	3600
TGGTAAAACT	TGGCAATGTT	ACCAATGATG	GTGATTTAAA	CATTACCACT	CACGCTAAAC	3660
GCAACCAAAG	AAGCATCATC	GGCGGAGATA	TAATCAACAA	AAAAGGAAGC	TTAAATATTA	3720
CAGACAGTAA	TAATGATGCT	GAAATCCAAA	TTGGCGGCAA	TATCTCGCAA	AAAGAAGGCA	3780
ACCTCACGAT	TTCTTCCGAT	ATAATTAAAA	TCACCAAACA	GATAACAATC	AAAAAGGGTA	3840
TTGATGGAGA	GGACTCTAGT	TCAGATGCGA	CAAGTAATGC	CAACCTAACT	ATTAAAACCA	3900
AAGAATTGAA	ATTGACAGAA	GACCTAAGTA	TTTCAGGTTT	CAATAAAGCA	GAGATTACAG	3960
CCAAAGATGG	TAGAGATTTA	ACTATTGGCA	ACAGTAATGA	CGGTAACAGC	GGTGCCGAAG	4020

CCAAAACAGT	AACTTTTAAC	AATGTTAAAG	ATTCAAAAAT	CTCTGCTGAC	GGTCACAATG	408
TGACACTAAA	TAGCAAAGTG	AAAACATCTA	GCAGCAATGG	CGGACGTGAA	AGCAATAGCG	4140
ACAACGATAC	CGGCTTAACT	ATTACTGCAA	AAAATGTAGA	AGTAAACAAA	GATATTACTT	4200
CTCTCAAAAC	AGTAAATATC	ACCGCGTCGG	AAAAGGTTAC	CACCACAGCA	GGCTCGACCA	4260
TTAACGCAAC	AAATGGCAAA	GCAAGTATTA	CAACCAAAAC	AGGTGATATC	AGCGGTACGA	4320
TTTCCGGTAA	CACGGTAAGT	GTTAGCGCGA	CTGGTGATTT	AACCACTAAA	TCCGGCTCAA	4380
AAATTGAAGC	GAAATCGGGT	GAGGCTAATG	TAACAAGTGC	AACAGGTACA	ATTGGCGGTA	4440
CAATTTCCGG	TAATACGGTA	AATGTTACGG	CAAACGCTGG	CGATTTAACA	GTTGGGAATG	4500
GCGCAGAAAT	TAATGCGACA	GAAGGAGCTG	CAACCTTAAC	CGCAACAGGG	AATACCTTGA	4560
CTACTGAAGC	CGGTTCTAGC	ATCACTTCAA	CTAAGGGTCA	GGTAGACCTC	TTGGCTCAGA	4620
ATGGTAGCAT	CGCAGGAAGC	ATTAATGCTG	CTAATGTGAC	ATTAAATACT	ACAGGCACCT	4680
TAACCACCGT	GGCAGGCTCG	GATATTAAAG	CAACCAGCGG	CACCTTGGTT	ATTAACGCAA	4740
AAGATGCTAA	GCTAAATGGT	GATGCATCAG	GTGATAGTAC	AGAAGTGAAT	GCAGTCAACG	4800
ACTGGGGATT	TGGTAGTGTG	ACTGCGGCAA	CCTCAAGCAG	TGTGAATATC	ACTGGGGATT	4860
TAAACACAGT	AAATGGGTTA	AATATCATTT	CGAAAGATGG	TAGAAACACT	GTGCGCTTAA	4920
GAGGCAAGGA	AATTGAGGTG	AAATATATCC	AGCCAGGTGT	AGCAAGTGTA	GAAGAAGTAA	4980
TTGAAGCGAA	ACGCGTCCTT	Gaaaaagtaa	AAGATTTATC	TGATGAAGAA	AGAGAAACAT	5040
TAGCTAAACT	TGGTGTAAGT	GCTGTACGTT	TTGTTGAGCC	AAATAATACA	ATTACAGTCA	5100
ATACACAAAA	TGAATTTACA	ACCAGACCGT	CAAGTCAAGT	GATAATTTCT	GAAGGTAAGG	5160
CGTGTTTCTC	AAGTGGTAAT	GGCGCACGAG	TATGTACCAA	TGTTGCTGAC	GATGGACAGC	5220
CGTAGTCAGT	AATTGACAAG	GTAGATTTCA	TCCTGCAATG	AAGTCATTTT	ATTTTCGTAT	5280
TATTTACTGT	GTGGGTTAAA	GTTCAGTACG	GGCTTTACCC	ATCTTGTAAA	AAATTACGGA	5340
GAATACAATA	AAGTATTTTT	AACAGGTTAT	TATTATGAAA	AAAAATATAA	GCAGATTAAA	5400
ACTCAGTGCA	ATATCAGTAT	TGCTTGGCCT	GGCTTCTTCA	TCATTGTATG	CAGAAGAAGC	5460
GTTTTTAGTA	AAAGGCTTTC	AGTTATCTGG	TGCACTTGAA	ACTTTAAGTG	AAGACGCCCA	5520
ACTGTCTGTA	GCAAAATCTT	TATCTAAATA	CCAAGGCTCG	CAAACTTTAA	CAAACCTAAA	5580
AACAGCACAG	CTTGAATTAC	AGGCTGTGCT	AGATAAGATT	GAGCCAAATA	AATTTGATGT	5640
GATATTGCCG	CAACAAACCA	TTACGGATGG	CAATATCATG	TTTGAGCTAG	TCTCGAAATC	5700
AGCCGCAGAA	AGCCAAGTTT	TTTATAAGGC	GAGCCAGGGT	TATAGTGAAG	AAAATATCGC	5760
TCGTAGCCTG	CCATCTTTGA	AACAAGGAAA	AGTGTATGAA	GATGGTCGTC	AGTGGTTCGA	5820
TTTGCGTGAA	TTTAATATGG	CAAAAGAAAA	CCCGCTTAAG	GTTACCCGTG	TACATTACGA	5880
ACTAAACCCT	AAAAACAAAA	CCTCTAATTT	GATAATTGCG	GGCTTCTCGC	CTTTTGGTAA	5940
AACGCGTAGC	TTTATTTCTT	ATGATAATTT	CGGCGCGAGA	GAGTTTAACT	ACCAACGTGT	6000
AAGCTTGGGT	TTTGTTAATG	CCAATTTAAC	TGGTCATGAT	GATGTGTTAA	TTATACCAGT	6060

ATGAGTTATG	CTGATTCTAA	TGATATCGAC	GGCTTACCAA	GTGCGATTAA	TCGTAAATTA	6120
TCAAAAGGTC	AATCTATCTC	TGCGAATCTG	AAATGGAGTT	ATTATCTCCC	AACATTTAAC	6180
CTTGGCATGG	AAGACCAATT	TAAAATTAAT	TTAGGCTACA	ACTACCGCCA	TATTAATCAA	6240
ACCTCCGCGT	TAAATCGCTT	GGGTGAAACG	AAGAAAAAAT	TTGCAGTATC	AGGCGTAAGT	6300
GCAGGCATTG	ATGGACATAT	CCAATTTACC	CCTAAAACAA	TCTTTAATAT	TGATTTAACT	6360
CATCATTATT	ACGCGAGTAA	ATTACCAGGC	TCTTTTGGAA	TGGAGCGCAT	TGGCGAAACA	6420
TTTAATCGCA	GCTATCACAT	TAGCACAGCC	AGTTTAGGGT	TGAGTCAAGA	GTTTGCTCAA	6480
GGTTGGCATT	TTAGCAGTCA	ATTATCAGGT	CAATTTACTC	TACAAGATAT	TAGCAGTATA	6540
GATTTATTCT	CTGTAACAGG	TACTTATGGC	GTCAGAGGCT	TTAAATACGG	CGGTGCAAGT	6600
GGTGAGCGCG	GTCTTGTATG	GCGTAATGAA	TTAAGTATGC	CAAAATACAC	CCGCTTCCAA	6660
ATCAGCCCTT	ATGCGTTTTA	TGATGCAGGT	CAGTTCCGTT	ATAATAGCGA	AAATGCTAAA	6720
ACTTACGGCG	AAGATATGCA	CACGGTATCC	TCTGCGGGTT	TAGGCATTAA	AACCTCTCCT	6780
ACACAAAACT	TAAGCCTAGA	TGCTTTTGTT	GCTCGTCGCT	TTGCAAATGC	CAATAGTGAC	6840
aatttgaatg	GCAACAAAAA	ACGCACAAGC	TCACCTACAA	CCTTCTGGGG	GAGATTAACA	6900
TTCAGTTTCT	AACCCTGAAA	TTTAATCAAC	TGGTAAGCGT	TCCGCCTACC	AGTTTATAAC	6960
TATATGCTTT	ACCCGCCAAT	TTACAGTCTA	TAGGCAACCC	TGTTTTTACC	CTTATATATC	7020
AAATAAACAA	GCTAAGCTGA	GCTAAGCAAA	CCAAGCAAAC	TCAAGCAAGC	CAAGTAATAC	7080
TAAAAAAACA	ATTTATATGA	TAAACTAAAG	TATACTCCAT	GCCATGGCGA	TACAAGGGAT	7140
TATAATAT	GACAAAAGAA	AATTTGCAAA	ACGCTCCTCA	AGATGCGACC	GCTTTACTTG	7200
CGGAATTAAG	CAACAATCAA	ACTCCCCTGC	GAATATTTAA	ACAACCACGC	AAGCCCAGCC	7260
PATTACGCTT	GGAACAACAT	ATCGCAAAAA	AAGATTATGA	GTTTGCTTGT	CGTGAATTAA	7320
IGGTGATTCT	GGAAAAAATG	GACGCTAATT	TTGGAGGCGT	TCACGATATT	GAATTTGACG	7380
CACCCGCTCA	GCTGGCATAT	CTACCCGAAA	AATTACTAAT	TTATTTTGCC	ACTCGTCTCG	7440
CTAATGCAAT	TACAACACTC	TTTTCCGACC	CCGAATTGGC	AATTTCTGAA	GAAGGGGCGT	7500
TAAAGATGAT	TAGCCTGCAA	CGCTGGTTGA	CGCTGATTTT	TGCCTCTTCC	CCCTACGTTA	7560
ACGCAGACCA	TATTCTCAAT	ATAATATAAA	TCAACCCAGA	TTCCGAAGGT	GGCTTTCATT	7620
ragcaacaga	CAACTCTTCT	ATTGCTAAAT	TCTGTATTTT	TTACTTACCC	GAATCCAATG	7680
CAATATGAG	TTTAGATGCG	TTATGGGCAG	GGAATCAACA	ACTTTGTGCT	TCATTGTGTT	7740
TTGCGTTGCA	GTCTTCACGT	TTTATTGGTA	CCGCATCTGC	GTTTCATAAA	AGAGCGGTGG	7800
				•	TTGCCTGCAA	7860
					AAGCACGATG	7920
	•				TGGCAAGACC	7980
					CTGCTTGAAC	8040
TTTTAATTC	GGGACATTCG	ATTTATCGTA	CACATTCAAC	TTCAATGATT	GCTGCTCGAG	8100

AAAAATTCTA	TTTAGTCGGC	TTAGGCCATG	AGGGCGTTGA	TAAAATAGGT	CGAGAAGTGT	8160
TTGACGAGTT	CTTTGAAATC	AGTAGCAATA	ATATAATGGA	GAGACTGTTT	TTTATCCGTA	8220
AACAGTGCGA	AACTTTCCAA	CCCGCAGTGT	TCTATATGCC	AAGCATTGGC	ATGGATATTA	8280
CCACGATTTT	TGTGAGCAAC	ACTCGGCTTG	CCCCTATTCA	AGCTGTAGCC	CTGGGTCATC	8340
CTGCCACTAC	GCATTCTGAA	TTTATTGATT	ATGTCATCGT	AGAAGATGAT	TATGTGGGCA	8400
GTGAAGATTG	TTTCAGCGAA	ACCCTTTTAC	GCTTACCCAA	AGATGCCCTA	CCTTATGTAC	8460
CTTCTGCACT	CGCCCCACAA	AAAGTGGATT	ATGTACTCAG	GGAAAACCCT	GAAGTAGTCA	8520
ATATCGGTAT	TGCCGCTACC	ACAATGAAAT	TAAACCCTGA	ATTTTTGCTA	ACATTGCAAG	8580
AAATCAGAGA	TAAAGCTAAA	GTCAAAATAC	ATTTTCATTT	CGCACTTGGA	CAATCAACAG	8640
GCTTGACACA	CCCTTATGTC	AAATGGTTTA	TCGAAAGCTA	TTTAGGTGAC	GATGCCACTG	8700
CACATCCCCA	CGCACCTTAT	CACGATTATC	TGGCAATATT	GCGTGATTGC	GATATGCTAC	8760
TAAATCCGTT	TCCTTTCGGT	AATACTAACG	GCATAATTGA	TATGGTTACA	TTAGGTTTAG	8820
TTGGTGTATG	CAAAACGGGG	GATGAAGTAC	ATGAACATAT	TGATGAAGGT	CTGTTTAAAC	8880
GCTTAGGACT	ACCAGAATGG	CTGATAGCCG	ACACACGAGA	AACATATATT	GAATGTGCTT	8940
TGCGTCTAGC	AGAAAACCAT	CAAGAACGCC	TTGAACTCCG	TCGTTACATC	ATAGAAAACA	9000
ACGGCTTACA	AAAGCTTTTT	ACAGGCGACC	CTCGTCCATT	GGGCAAAATA	CTGCTTAAGA	9060
AAACAAATGA	ATGGAAGCGG	AAGCACTTGA	GTAAAAAATA	ACGGTTTTTT	AAAGTAAAAG	9120
<b>IGCGGTTAAT</b>	TTTCAAAGCG	TTTTAAAAAC	CTCTCAAAAA	TCAACCGCAC	TTTTATCTTT	9180
ATAACGATCC	CGCACGCTGA	CAGTTTATCA	GCCTCCCGCC	ATAAAACTCC	GCCTTTCATG	9240
ECGGAGATTT	TAGCCAAAAC	TGGCAGAAAT	TAAAGGCTAA	AATCACCAAA	TTGCACCACA	9300
AAATCACCAA	TACCCACAAA	AAA				9323

### (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4287 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: DNA (genomic)

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCAATCTG	GGCGATATTT	TTGCCAAAGG	TGGTAACATT	AATGTCCGCG	CTGCCACTAT	60
TCGCAATAAA	GGTAAACTTT	CTGCCGACTC	TGTAAGCAAA	GATAAAAGTG	GTAACATTGT	120
TCTCTCTGCC	AAAGAAGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTTCCGCTC	AAAATCAGCA	180
AGCCAAAGGT	GGTAAGTTGA	TGATTACAGG	CGATAAAGTT	ACATTGAAAA	CGGGTGCACT	240
TATCGACCTT	TCGGGTAAAG	AAGGGGGAGA	AACTTATCTT	GGCGGTGACG	AGCGTGGCGA	300
AGGTAAAAAC	GGCATTCAAT	TAGCAAAGAA	AACCACTTTA	GAAAAAGGCT	CAACAATTAA	360

IGIGICAGGI	AAAGAAAAG	CIGGGCGCGC	INTIGIATEG	GGCGATATTG	CGTTAATTGA	42
CGGCAATATT	AATGCCCAAG	GTAAAGATAT	CGCTAAAACT	GGTGGTTTTG	TGGAGACGTC	48
GGGGCATTAC	TTATCCATTG	ATGATAACGC	AATTGTTAAA	ACAAAAGAAT	GGCTACTAGA	54
CCCAGAGAAT	GTGACTATTG	AAGCTCCTTC	CGCTTCTCGC	GTCGAGCTGG	GTGCCGATAG	60
GAATTCCCAC	TCGGCAGAGG	TGATAAAAGT	GACCCTAAAA	AAAAATAACA	CCTCCTTGAC	66
AACACTAACC	AATACAACCA	TTTCAAATCT	TCTGAAAAGT	GCCCACGTGG	TGAACATAAC	72
GGCAAGGAGA	AAACTTACCG	TTAATAGCTC	TATCAGTATA	GAAAGAGGCT	CCCACTTAAT	78
TCTCCACAGT	GAAGGTCAGG	GCGGTCAAGG	TGTTCAGATT	GATAAAGATA	TTACTTCTGA	84
AGGCGGAAAT	TTAACCATTT	ATTCTGGCGG	ATGGGTTGAT	GTTCATAAAA	ATATTACGCT	90
TGGTAGCGGC	TTTTTAAACA	TCACAACTAA	AGAAGGAGAT	ATCGCCTTCG	AAGACAAGTC	96
TGGACGGAAC	AACCTAACCA	TTACAGCCCA	AGGGACCATC	ACCTCAGGTA	ATAGTAACGG	102
CTTTAGATTT	AACAACGTCT	CTCTAAACAG	CCTTGGCGGA	AAGCTGAGCT	TTACTGACAG	1080
CAGAGAGGAC	AGAGGTAGAA	GAACTAAGGG	TAATATCTCA	AACAAATTTG	ACGGAACGTT	1140
AAACATTTCC	GGAACTGTAG	ATATCTCAAT	GAAAGCACCC	AAAGTCAGCT	GGTTTTACAG	1200
agacaaagga	CGCACCTACT	GGAACGTAAC	CACTTTAAAT	GTTACCTCGG	GTAGTAAATT	1260
TAACCTCTCC	ATTGACAGCA	CAGGAAGTGG	CTCAACAGGT	CCAAGCATAC	GCAATGCAGA	1320
ATTAAATGGC	ATAACATTTA	ATAAAGCCAC	TTTTAATATC	GCACAAGGCT	CAACAGCTAA	1380
CTTTAGCATC	AAGGCATCAA	TAATGCCCTT	TAAGAGTAAC	GCTAACTACG	CATTATTTAA	1440
TGAAGATATT	TCAGTCTCAG	GGGGGGTAG	CGTTAATTTC	AAACTTAACG	CCTCATCTAG	1500
CAACATACAA	ACCCCTGGCG	TAATTATAAA	ATCTCAAAAC	TTTAATGTCT	CAGGAGGGTC	1560
AACTTTAAAT	CTCAAGGCTG	AAGGTTCAAC	AGAAACCGCT	TTTTCAATAG	AAAATGATTT	1620
AAACTTAAAC	GCCACCGGTG	GCAATATAAC	AATCAGACAA	GTCGAGGGTA	CCGATTCACG	1680
CGTCAACAAA	GGTGTCGCAG	CCAAAAAAA	CATAACTTTT	AAAGGGGGTA	ATATCACCTT	1740
CGGCTCTCAA	AAAGCCACAA	CAGAAATCAA	AGGCAATGTT	ACCATCAATA	AAAACACTAA	1800
CGCTACTCTT	CGTGGTGCGA	ATTTTGCCGA	AAACAAATCG	CCTTTAAATA	TAGCAGGAAA	1860
IGTTATTAAT	AATGGCAACC	TTACCACTGC	CGGCTCCATT	ATCAATATAG	CCGGAAATCT	1920
TACTGTTTCA	AAAGGCGCTA	ACCTTCAAGC	TATAACAAAT	TACACTTTTA	ATGTAGCCGG	1980
CTCATTTGAC	AACAATGGCG	CTTCAAACAT	TTCCATTGCC	AGAGGAGGGG	CTAAATTTAA	2040
AGATATCAAT	AACACCAGTA	GCTTAAATAT	TACCACCAAC	TCTGATACCA	CTTACCGCAC	2100
CATTATAAAA	GGCAATATAT	CCAACAAATC	AGGTGATTTG	AATATTATTG	ATAAAAAAG	2160
CGACGCTGAA	ATCCAAATTG	GCGGCAATAT	CTCACAAAAA	GAAGGCAATC	TCACAATTTC	2220
TTCTGATAAA	GTAAATATTA	CCAATCAGAT	AACAATCAAA	GCAGGCGTTG	AAGGGGGGCG	2280
ITCTGATTCA	AGTGAGGCAG	AAAATGCTAA	CCTAACTATT	CAAACCAAAG	AGTTAAAATT	2340
GGCAGGAGAC	CTAAATATTT	CAGGCTTTAA	TAAAGCAGAA	ATTACAGCTA	AAAATGGCAG	2400

TGATTTAACT	ATTGGCAATG	CTAGCGGTGG	TAATGCTGAT	GCTAAAAAA	TGACTTTTGA	2460
CAAGGTTAA	GATTCAAAAA	. TCTCGACTGA	CGGTCACAAT	GTAACACTAA	ATAGCGAAGT	2520
GAAAACGTCT	AATGGTAGTA	GCAATGCTGG	TAATGATAAC	AGCACCGGTT	TAACCATTTC	2580
CGCAAAAGAT	GTAACGGTAA	ACAATAACGT	TACCTCCCAC	AAGACAATAA	ATATCTCTGC	2640
CGCAGCAGGA	AATGTAACAA	CCAAAGAAGG	CACAACTATC	AATGCAACCA	CAGGCAGCGT	2700
GGAAGTAACT	GCTCAAAATG	GTACAATTAA	AGGCAACATT	ACCTCGCAAA	ATGTAACAGT	2760
GACAGCAACA	GAAAATCTTG	TTACCACAGA	GAATGCTGTC	ATTAATGCAA	CCAGCGGCAC	2820
AGTAAACATT	AGTACAAAAA	CAGGGGATAT	TAAAGGTGGA	ATTGAATCAA	CTTCCGGTAA	2880
TGTAAATATT	ACAGCGAGCG	GCAATACACT	TAAGGTAAGT	AATATCACTG	GTCAAGATGT	2940
AACAGTAACA	GCGGATGCAG	GAGCCTTGAC	AACTACAGCA	GGCTCAACCA	TTAGTGCGAC	3000
AACAGGCAAT	GCAAATATTA	CAACCAAAAC	AGGTGATATC	AACGGTAAAG	TTGAATCCAG	3060
CTCCGGCTCT	GTAACACTTG	TTGCAACTGG	AGCAACTCTT	GCTGTAGGTA	ATATTTCAGG	3120
TAACACTGTT	ACTATTACTG	CGGATAGCGG	TAAATTAACC	TCCACAGTAG	GTTCTACAAT	3180
TAATGGGACT	AATAGTGTAA	CCACCTCAAG	CCAATCAGGC	GATATTGAAG	GTACAATTTC	3240
TGGTAATACA	GTAAATGTTA	CAGCAAGCAC	TGGTGATTTA	ACTATTGGAA	ATAGTGCAAA	3300
AGTTGAAGCG	AAAAATGGAG	CTGCAACCTT	AACTGCTGAA	TCAGGCAAAT	TAACCACCCA	3360
AACAGGCTCT	AGCATTACCT	CAAGCAATGG	TCAGACAACT	CTTACAGCCA	AGGATAGCAG	3420
TATCGCAGGA	AACATTAATG	CTGCTAATGT	GACGTTAAAT	ACCACAGGCA	CTTTAACTAC	3480
TACAGGGGAT	TCAAAGATTA	ACGCAACCAG	TGGTACCTTA	ACAATCAATG	CAAAAGATGC	3540
CAAATTAGAT	GGTGCTGCAT	CAGGTGACCG	CACAGTAGTA	AATGCAACTA	ACGCAAGTGG	3600
CTCTGGTAAC	GTGACTGCGA	AAACCTCAAG	CAGCGTGAAT	ATCACCGGGG	ATTTAAACAC	3660
AATAAATGGG	TTAAATATCA	TTTCGGAAAA	TGGTAGAAAC	ACTGTGCGCT	TAAGAGGCAA	3720
GGAAATTGAT	GTGAAATATA	TCCAACCAGG	TGTAGCAAGC	GTAGAAGAGG	TAATTGAAGC	3780
GAAACGCGTC	CTTGAGAAGG	TAAAAGATTT	ATCTGATGAA	GAAAGAGAAA	CACTAGCCAA	3840
ACTTGGTGTA	AGTGCTGTAC	GTTTCGTTGA	GCCAAATAAT	GCCATTACGG	TTAATACACA	3900
					AGGCGTGTTT	3960
CTCAAGTGGT	AATGGCGCAC	GAGTATGTAC	CAATGTTGCT	GACGATGGAC	AGCAGTAGTC	4020
					TATTATTTAC	4080
					GAAAAATACA	4140
					AAAACTCAGT	4200
			TCATCGACGT	ATGCAGAAGA	AGCGTTTTTA	4260
GTAAAAGGCT	TTCAGTTATC	TGGCGCG				4287

### (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4702 base pairs (B) TYPE: nucleic acid

  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

60	AAACCACTAT	GACGGCAATA	CATGCAAGTA	GTACAGCAAC	GTCGTACACG	GGGAATGAGC
120	AAAATGAAAT	AACATTGACC	GAAACAATTT	TCATCAATTG	ATCAATGCTA	CCGTAATAGC
180	CATCTGACCA	AACCGTGTTA	TGCCGTTTTC	GCAGCAACTC	TTACAAGAAA	GGAGCAGTTT
240	TCAACCCAAA	GTCTTTTTAA	TAACGGACAA	TTTTAGATTC	TTAAAAGGGA	AATCTCCCAA
300	CTTCTACGCT	GGCTTTACTG	TAACACTAAT	ACGCAATTAT	ATAGGTAAAG	TGGTATCACA
360	CCAAGGATAA	CTTGAGCAAA	TAATTTCACC	TCAAGGCGCG	AACGAAAACA	AGACATTTCT
420	GTAGCGTAAA	GGTAAAGACG	AATTACCGTT	ATCACGGTTT	GAAATCGTGA	AGCACTCGCT
480	GTAGTATTTC	GTAAATGGCG	CGTGATTAGC	AAAACGAGGG	GGCAAAGTĠA	CCTTATTGGT
540	TCACTTACAG	AATCCAACCA	CGATATAATA	TCACCATCAG	GGGCAAAAA	TTTACTTGCA
600	aaggtggtaa	ATTTTTGCCA	TCTGGGCGAT	AAGCGATCAA	CCTGAAAACG	CATTGCTGCA
660	ACTCTGTAAG	CTTTCTGCCG	TAAAGGTAAA	CTATTCGCAA	CGCGCTGCCA	CATTAATGTC
720	AAATTGGCGG	GGTGAAGCGG	TGCCAAAGAA	TTGTTCTCTC	AGTGGTAACA	CAAAGATAAA
780	CAGGTGATAA	TTGATGATTA	AGGTGGTAAG	AGCAAGCCAA	GCTCAAAATC	TGTAATTTCC
840	GAGAGACTTA	AAAGAAGGGG	CCTTTCAGGT	CAGTTATCGA	AAAACAGGTG	AGTCACATTA
900	AGAAAACCTC	CAATTAGCGA	AAATGGTATT	GCGAAGGTAA	GATGAGCGTG	TCTTGGCGGT
960	GCGCTATTGT	AAAGGCGGGC	AGGCAAAGAA	TTAATGTATC	GGCTCGACAA	TTTAGAAAAA
1020	ATATTGCTAA	CAAGGTAGCG	CATTAATGCT	TTAATGGTAA	ATTGCATTAA	ATGGGGCGAT
1080	ATGTGATTGT	ATTGGTGATG	TGACTTATCC	CATCAGGACA	TTTGTGGAAA	AACTGGCGGC
1140	TTACATCTGG	ATTGÄAACTC	TGATGTGTCC	TÄGACCCAGA	GAGTGGTTAT	TGACGCTAAA
1200	AAGAGTCACC	GATGGGACTA	TACAACAGGA	ACCAAGGATA	ACCGGCGAAA	ACGCAATAAT
1260	AAATCCTAAG	ACTCTTGAGC	AACAAACTCA	AACCTACATT	AGTATTTCTA	TAAAGGTAAT
1320	GCTCCATCAA	TATGTTAATA	TAATAGAATT	TCACTGCTAA	TATGTTAATA	AAGAGGTTCT
1380	TTAACGGTGA	GGAGTTAAAA	TAAACGAGAT	CACTTCACAC	GGCAGTTTAA	CTTATCTAAT
1440	TTGATGTTCA	GGCTCTTGGG	CATTAAAGCA	GTAATTTAAC	AACGAAAATG	TATTACCTCA
1500	CTGTAGCTTT	GCTGGGGATT	CAATATTGTC	CGGGTTTTTT	ACGCTTGGTA	TAAAAACATC
1560	CACAAGGGAC	CAAATTACCG	AACAGATGCT	CACGTAACGC	GGCGATAAAG	TGAGAGAGAG
1620	TTAACGGGAC	AATGTATCTA	TAGATTCAAT	ATAAACAATT	AATAAAGATG	GATAACCGTC

GGGCAAGGGT	TTAAAGTTTA	TTGCAAATCA	AAATAATTTC	ACTCATAAAT	TTGATGGCGA	1680
AATTAACATA	TCTGGAATAG	TAACAATTAA	CCAAACCACG	AAAAAAGATG	TTAAATACTG	1740
GAATGCATCA	AAAGACTCTT	ACTGGAATGT	TTCTTCTCTT	ACTTTGAATA	CGGTGCAAAA	1800
ATTTACCTTT	ATAAAATTCG	TTGATAGCGG	CTCAAATTCC	CAAGATTTGA	GGTCATCACG	1860
TAGAAGTTTT	GCAGGCGTAC	ATTTTAACGG	CATCGGAGGC	AAAACAAACT	TCAACATCGG	1920
AGCTAACGCA	AAAGCCTTAT	TTAAATTAAA	ACCAAACGCC	GCTACAGACC	CAAAAAAAGA	1980
ATTACCTATT	ACTTTTAACG	CCAACATTAC	AGCTACCGGT	AACAGTGATA	GCTCTGTGAT	2040
GTTTGACATA	CACGCCAATC	TTACCTCTAG	AGCTGCCGGC	ATAAACATGG	ATTCAATTAA	2100
CATTACCGGC	GGGCTTGACT	TTTCCATAAC	ATCCCATAAT	CGCAATAGTA	ATGCTTTTGA	2160
AATCAAAAAA	GACTTAACTA	TAAATGCAAC	TGGCTCGAAT	TTTAGTCTTA	AGCAAACGAA	2220
AGATTCTTTT	TATAATGAAT	ACAGCAAACA	CGCCATTAAC	TCAAGTCATA	ATCTAACCAT	2280
TCTTGGCGGC	AATGTCACTC	TAGGTGGGGA	AAATTCAAGC	AGTAGCATTA	CGGGCAATAT	2340
CAATATCACC	AATAAAGCAA	ATGTTACATT	ACAAGCTGAC	ACCAGCAACA	GCAACACAGG	2400
CTTGAAGAAA	AGAACTCTAA	CTCTTGGCAA	TATATCTGTT	GAGGGGAATT	TAAGCCTAAC	2460
TGGTGCAAAT	GCAAACATTG	TCGGCAATCT	TTCTATTGCA	GAAGATTCCA	CATTTAAAGG	2520
AGAAGCCAGT	GACAACCTAA	ACATCACCGG	CACCTTTACC	AACAACGGTA	CCGCCAACAT	2580
TAATATAAAA	CAAGGAGTGG	TAAAACTCCA	AGGCGATATT	ATCAATAAAG	GTGGTTTAAA	2640
TATCACTACT	AACGCCTCAG	GCACTCAAAA	AACCATTATT	AACGGAAATA	TAACTAACGA	2700
AAAAGGCGAC	TTAAACATCA	AGAATATTAA	AGCCGACGCC	GAAATCCAAA	TTGGCGGCAA	2760
TATCTCACAA	AAAGAAGGCA	ATCTCACAAT	TTCTTCTGAT	AAAGTAAATA	TTACCAATCA	2820
GATAACAATC	AAAGCAGGCG	TTGAAGGGGG	GCGTTCTGAT	TCAAGTGAGG	CAGAAAATGC	2880
TAACCTAACT	ATTCAAACCA	AAGAGTTAAA	ATTGGCAGGA	GACCTAAATA	TTTCAGGCTT	2940
TAATAAAGCA	GAAATTACAG	CTAAAAATGG	CAGTGATTTA	ACTATTGGCA	ATGCTAGCGG	3000
TGGTAATGCT	GATGCTAAAA	AAGTGACTTT	TGACAAGGTT	AAAGATTCAA	AAATCTCGAC	3060
TGACGGTCAC	AATGTAACAC	TAAATAGCGA	AGTGAAAACG	TCTAATGGTA	GTAGCAATGC	3120
TGGTAATGAT	AACAGCACCG	GTTTAACCAT	TTCCGCAAAA	GATGTAACGG	TAAACAATAA	3180
CGTTACCTCC	CACAAGACAA	TAAATATCTC	TGCCGCAGCA	GGAAATGTAA	CAACCAAAGA	3240
AGGCACAACT	ATCAATGCAA	CCACAGGCAG	CGTGGAAGTA	ACTGCTCAAA	ATGGTACAAT	3300
TAAAGGCAAC	ATTACCTCGC	AAAATGTAAC	AGTGACAGCA	ACAGAAAATC	TTGTTACCAC	3360
AGAGAATGCT	GTCATTAATG	CAACCAGCGG	CACAGTAAAC	ATTAGTACAA	AAACAGGGGA	3420
TATTAAAGGT	GGAATTGAAT	CAACTTCCGG	TAATGTAAAT	ATTACAGCGA	GCGGCAATAC	3480
ACTTAAGGTA	AGTAATATCA	CTGGTCAAGA	TGTAACAGTA	ACAGCGGATG	CAGGAGCCTT	3540
GACAACTACA	GCAGGCTCAA	CCATTAGTGC	GACAACAGGC	AATGCAAATA	TTACAACCAA	3600
AACAGGTGAT	ATCAACGGTA	AAGTTGAATC	CAGCTCCGGC	TCTGTAACAC	TTGTTGCAAC	3660

TGGAGCAACT	CTTGCTGTAG	GTAATATTTC	AGGTAACACT	GTTACTATTA	CTGCGGATAG	372
CGGTAAATTA	ACCTCCACAG	TAGGTTCTAC	AATTAATGGG	ACTAATAGTG	TAACCACCTC	3780
AAGCCAATCA	GGCGATATTG	AAGGTACAAT	TTCTGGTAAT	ACAGTAAATG	TTACAGCAAG	384
CACTGGTGAT	TTAACTATTG	GAAATAGTGC	AAAAGTTGAA	GCGAAAAATG	GAGCTGCAAC	3900
CTTAACTGCT	GAATCAGGCA	AATTAACCAC	CCAAACAGGC	TCTAGCATTA	CCTCAAGCAA	3960
TGGTCAGACA	ACTCTTACAG	CCAAGGATAG	CAGTATCGCA	GGAAACATTA	ATGCTGCTAA	4020
TGTGACGTTA	AATACCACAG	GCACTTTAAC	TACTACAGGG	GATTCAAAGA	TTAACGCAAC	4080
CAGTGGTACC	TTAACAATCA	ATGCAAAAGA	TGCCAAATTA	GATGGTGCTG	CATCAGGTGA	4140
CCGCACAGTA	GTAAATGCAA	CTAACGCAAG	TGGCTCTGGT	AACGTGACTG	CGAAAACCTC	4200
AAGCAGCGTG	AATATCACCG	GGGATTTAAA	CACAATAAAT	GGGTTAAATA	TCATTTCGGA	4260
AAATGGTAGA	AACACTGTGC	GCTTAAGAGG	CAAGGAAATT	GATGTGAAAT	ATATCCAACC	4320
AGGTGTAGCA	AGCGTAGAAG	aggtaattga	AGCGAAACGC	GTCCTTGAGA	AGGTAAAAGA	4380
TTTATCTGAT	GAAGAAAGAG	AAACACTAGC	CAAACTTGGT	GTAAGTGCTG	TACGTTTCGT	4440
TGAGCCAAAT	AATGCCATTA	CGGTTAATAC	ACAAAACGAG	TTTACAACCA	AACCATCAAG	4500
TCAAGTGACA	ATTTCTGAAG	GTAAGGCGTG	TTTCTCAAGT	GGTAATGGCG	CACGAGTATG	4560
TACCAATGTT	GCTGACGATG	GACAGCAGTA	GTCAGTAATT	GACAAGGTAG	ATTTCATCCT	4620
GCAATGAAGT	CATTTTATTT	TCGTATTATT	TACTGTGTGG	GTTAAAGTTC	AGTACGGGCT	4680
TTACCCACCT	TGTAAAAAAT	TA				4702

58

### CLAIMS

What we claim is:

- 1. A vaccin against disease caused by non-typeable Haemophilus influenzae, including otitis media, sinusitis and bronchitis, comprising an effective amount of a high molecular weight protein of non-typeable Haemophilus influenzae which is protein HMW1, HMW2, HMW3 or HMW4 or a variant or fragment of said protein retaining immunological properties thereof or a synthetic peptide having an amino acid sequence corresponding to that of said protein, and a physiological carrier therefor.
- 2. The vaccine of claim 1 wherein said protein is HMW1 encoded by the DNA sequence shown in Figure 1 (SEQ ID NO:1), having the derived amino acid sequence of Figure 2 (SEQ ID NO:2) and having an apparent molecular weight of 125 kDa.
- 3. The vaccine of claim 1 wherein said protein is HMW2 encoding by the DNA sequence shown in Figure 3 (SEQ ID NO:3), having the derived amino acid sequence of Figure 4 (SEQ ID NO:4) and having an apparent molecular weight of 120 kDa.

# PROTEIN HIGH MOLECULAR WEIGHT FIG.1A. DNA SEQUENCE OF (HMW1

ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACATGCCCTG ATGAACCGAG GGAAGGGAGG GAGGGCCAAG AATGAAGAGG GAGCTGAACG TTAAAAATA TCATCTTTCA TGCTGTGTCT GAATTGGCAC GGGGTTGTGA CCATTCCACA GAAAAAGGCA AGTTTTTACA CATCTTTCAT AGGAGAAAAT ATGCTTTGGT AAAGCCACTT AATCTGTTTT GCCACTATGC CGATATCATT AACCAAATCT GTATAAATCC GCCATATAAA ATGGTATAAT CTTTCATCTT AACTAACCTT CAAATTCAGC AAACGCCTGA ACTTAGCGTT TCTATTCCAC TAATAAAACC ATTATCCGCA ACAGTGTTGA TGTTACATCT TTCATCTTTC TTACAAGGAA TGGATGTAGT ACACGGCACA TTTCATCTTT GAAATGGTGC TGCAAATATT ACAATTACAA CACCTTTTTT GCAGTCTATA TTCATCTTTC ATCTTTCATC TCATCTTTCA TCTTTCATCT TTAATTGTTC TGCTCGCATG AAAGTGCGTC TACTATCTTT AGGTGTAACA AATTTAACAT CGACCAAAAT AGAAAACAAC AACTCCGCCG TATTCAACCG ATAAAGTAAT TATATCGTCT AATTGGAAAC TCTTTCATCT CTTTCATCTT AACGCAAATG ATGAACAAGC GCGAAAAACC TCCGCTATGT AGCAAGCGGC AAGTAGATGG 251 51 101 151 201 301 351 401 451 501 551 651 601 701

1/68

## FIG. 1B

						2/9									
TTAATCAAC	CTAATGGCTT	GCGCGTAATT	TGTGAATCAC	TTGGTGGCAA	ATTTCTTTAC	AACCATTACT	GCGATATTTT	CGAAACCAAG	CAATATTGTT	TTTCCGCTCA	GATAAAGTCA	AGGGGGAGAA	GCATTCAATT	GTATCAGGCA	GTTAATTGAC
GACAAGTCTT	TAAAGACGCA ATTATTAACA	TTTCTAACGA AAACATCAAG	TCGCTGAAAT	GTAAATCTTA	TGGTGGCAGC	TAATAAACCC AACCATTACT	GTCAATCTGG GCGATATTTT	TGCCACTATT	ATAAAAGCGG	GGCGGTGTAA	GATTACAGGC	CAGGTAAAGA AGGGGGAGAA	GGTAAAAAGG	AACCATCAAT	CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC
CCCAATTAAA AGGGATTTTA GATTCTAACG GACAAGTCTT		TTTCTAACGA	GCAAACCAAA GATAAAGCGC	CTGTCGGTAA AGACGGCAGT	TTAGCGTAAA	ATCAGCGATA	AAATGAAGCG	ATGTCCGTGC	GTAAGCAAAG		GCAAGCTGAT	ATCGACCTTT		AAAAAGGCTC	ATTGTGTGGG
AGGGA'I"I'I'IA	TCACAATAGG	ACGCTAGACA	GCAAACCAAA	CTGTCGGTAA	GAGGGTGTGA	AAAAATCACC	CCGCCCTGA AAATGAAGCG	GGTAACATTA	TGCTGATTCT	AAGAGGGTGA AGCGGAAATT	GCTAAAGGCG GCAAGCTGAT	AGGTGCAGTT	GCGGTGACGA GCGCGGCGAA	ACCTCTTTAG	CGGACGCGCT
CCCAATTAAA	CCAAATGGTA	TACGGCTTCT	TCACCTTCGA	GGTTTAATTA	AGTGAAAAAC	TCGCAGGGCA	TACAGCATTG	TGCCAAAGGC	GTAAACTTTC	CTTTCCGCCA	AAATCAGCAA	CATTAAAAAC	ACTTACCTTG	AGCAAAGAAA	AAGAAAAAGG
12 <i>/</i>	801	851	901	951	1001	1051	1101	1151	1201	1251	1301	1351	1401	1451	1501

## FIG. 1C

1551	GGCAATATTA	ACGCTCAAGG	ACGCTCAAGG TAGTGGTGAT	ATCGCTAAAA CCGGTGGTTT	CCGGTGGTTT	
1601	TGTGGAGACG	TCGGGGCATG		ATTTATTCAT CAAAGACAAT	GCAATTGTTG	
1651	ACGCCAAAGA	GTGGTTGTTA	GACCCGGATA	ATGTATCTAT	TAATGCAGAA	
1701	ACAGCAGGAC	GCAGCAATAC	TTCAGAAGAC	TTCAGAAGAC GATGAATACA	CGGGATCCGG	
1751	GAATAGTGCC	AGCACCCCAA	AGCACCCCAA AACGAAACAA AGAAAAGACA	AGAAAAGACA	ACATTAACAA	
1801	ACACAACTCT	TGAGAGTATA	CTAAAAAAAG	GTACCTTTGT	TAACATCACT	
1851	GCTAATCAAC	GCATCTATGT	CAATAGCTCC ATTAATTTAT	ATTAATTTAT	CCAATGGCAG	. 3
1901	CTTAACTCTT	TGGAGTGAGG	GTCGGAGCGG	TGGCGGCGTT	GAGATTAACA	/ 68
1951	ACGATATTAC	CACCGGTGAT	GATACCAGAG	GATACCAGAG GTGCAAACTT AACAATTTAC	AACAATTTAC	3
2001	TCAGGCGGCT	GGGTTGATGT	TCATAAAAT ATCTCACTCG	ATCTCACTCG	GGGCGCAAGG	
2051	TAACATAAAC	ATTACAGCTA	AACAAGATAT	CGCCTTTGAG	AAAGGAAGCA	
2101	ACCAAGTCAT	TACAGGTCAA	TACAGGTCAA GGGACTATTA CCTCAGGCAA TCAAAAAGGT	CCTCAGGCAA	TCAAAAAGGT	
2151	TTTAGATTTA	ATAATGTCTC	TCTAAACGGC ACTGGCAGCG		GACTGCAATT	
2201	CACCACTAAA	AGAACCAATA	AATACGCTAT	CACAAATAAA	TTTGAAGGGA	
2251	CTTTAAATAT	TTCAGGGAAA	TTCAGGGAAA GTGAACATCT	CAATGGTTTT	ACCTAAAAAT	
2301	GAAAGTGGAT	ATGATAAATT	CAAAGGACGC	ACTTACTGGA ATTTAACCTC	ATTTAACCTC	

## FIG. 1D

4TA	ZAA	\TT	AAT	ŀAT	AAA	1/68 4/68	ო ეტე	;AG	ιΑΤ	GA	j. G	SC CC	ĞŢ	AG	ည
AAACGGTATA	CAAGAGTCAA	AGTTTGAATT	GAGTGTT	GTGTAGT	AGATTTAAAA	TTTAACTTTA 0	GCACCGAI	TTTGAAGGAG	CGAAGGCAAT	CGGATTTTGA	ATTAATAGCG	AAATCTTA	CTTTTAATGT	ATTGCCAAAG	AAGCATCA
CTTACCCAGC CTTATAATTT	GAACGAAATG	TAAGTATTCT	CGGGAGGGG GAGTGTTGAT	CTCTAACGTC CAAACCCCCG GTGTAGTTAT	TTTCAACAGG GTCAAGTTTA	TAGAGAAAGA	CAAGTTGAAG GCACCGATGG	TAGCCAAAAA AAACATAACC	TAACAGAAAT	CTTATCGGTT	AGATGTCATC	ATATAGCCGG AAATCTTACC	ACAAATTTCA	AAATATTTCC	CTTTAAAGAC ATTGATAATT CCAAGAATTT AAGCATCACC
	CTTTAATGTT	TAGGGATAAA	ATTTCAGTTT		TTTCAACAGG	GGCTTCTCAA	AACACTTTTG	TAGCCAAAAA	CTTTGGCTCC AGGAAAGCCG TAACAGAAAT	TAACGTCACT	CTATTAAAAA	AATATTGTCA	CAAAGCTATC ACAAATTTCA		ATTGATAATT
TGCAGGCACA	TCATTCAACA AAGACACTAC	AAGGCACCAA	TAATGGAAAC	TCGCCTCATC	TACTTTAATG	AACAAAAACT	GAGGCAACAT	AAAGGCATTG	CTTTGGCTCC	ATAACAACGC	AAACCTTTAA	CGCTGGAGGC	ACGCTAATTT	TTTGACAACA AAGGCAATTC	CTTTAAAGAC
GAAGCGATAG	TCATTCAACA	CTTTGACATC	ACGCATCATT	TTCACACTTC	AAATTCTAAA	CTTCAGGCTC	AATGCCACCG	AATGATTGGT	GTAACATCAC	GTTACTATCA	CAACCATCAA	GCAACCTTAC	GTTGAAAGTA	AGGCGGCTTG	GAGGGGCTCG
2401	2451	2501	2551	2601	2651	2701	2751	2801	2851	2901	2951	3001	3051	3101	3151

## FIG. 1E

3201	ACCAACTCCA		CCGCACTATT	GCTCCACTTA CCGCACTATT ATAAGCGGCA ATATAACCAA	ATATAACCAA	
	TAAAAACGGT	GATTTAAATA	TTACGAACGA	TTACGAACGA AGGTAGTGAT ACTGAAATGC	ACTGAAATGC	
	AAATTGGCGG	CGATGTCTCG	CAAAAAGAAG	CAAAAAGAAG GTAATCTCAC	GATTTCTTCT	
	GACAAAATCA	ATATTACCAA	ATATTACCAA ACAGATAACA ATCAAGGCAG	ATCAAGGCAG	GTGTTGATGG	
3401	GGAGAATTCC	GATTCAGACG	GGAGAATTCC GATTCAGACG CGACAAACAA	TGCCAATCTA	ACCATTAAAA	
	CCAAAGAATT	GAAATTAACG	CAAGACCTAA	ATATTTCAGG	TTTCAATAAA	
3501	GCAGAGATTA	CAGCTAAAGA	TGGTAGTGAT	TTAACTATTG	GTAACACCAA U	_
3551	TAGTGCTGAT	GGTACTAATG	CCAAAAAAGT	AACCTTTAAC	CAGGTTAAAG ®	100
3601	ATTCAAAAT	CTCTGCTGAC	CTCTGCTGAC GGTCACAAGG	TGACACTACA CAGCAAAGTG	CAGCAAAGTG	
	GAAACATCCG	GTAGTAATAA	GTAGTAATAA CAACACTGAA	GATAGCAGTG	ACAATAATGC	
	CGGCTTAACT	ATCGATGCAA	ATCGATGCAA AAAATGTAAC	AGTAAACAAC AATATTACTT	AATATTACTT	
	CTCACAAAGC	AGTGAGCATC	TCTGCGACAA GTGGAGAAAT		TACCACTAAA	
	ACAGGTACAA	CCATTAACGC	CCATTAACGC AACCACTGGT	AACGTGGAGA	TAACCGCTCA	
	AACAGGTAGT	ATCCTAGGTG	GAATTGAGTC	CAGCTCTGGC	TCTGTAACAC	
	TTACTGCAAC	CGAGGGCGCT	CTTGCTGTAA	GCAATATTTC	GGGCAACACC	
	GTTACTGTTA	CTGCAAATAG	CGGTGCATTA	CTGCAAATAG CGGTGCATTA ACCACTTTGG CAGGCTCTAC	CAGGCTCTAC	

### FIG. 1F

TACACAAAAT	TTACAGTCGA	AATAATACAA	TATTGAGCCA	CTGTACGTTT	4801
GGAGTAAGTG	AGCTAAACTT	GAGAAGCGTT	GATGAAGAAA	AGATTTATCT	4751
AGAAGGTAAA	CGCATCCTTG	TGAAGCGAAA	ATGAAGTAAT	GCAAGCGTAG	4701
ACCGGGTATA	AATACATTCA	ATTGATGTGA	AGGCGTTAAA	TACTGTTAAA	4651
ATAAACACCG	AAAAAACGGT	ATATCATTTC	AATGGATTAA	AATCACAATA	4601
CTGGGGATTT	GTGAACATCA	CTCAAGCAGA	TCGCGACAAC	GGCAGCGTAA	4551
AAATGGCTCC	CAACCAACGC	GTGGTAAATG	TAACCACACA	CAGCATTGGG	4501
CTAAATGGCG	AGACGCTGAG	TTAACGCAAA	ACCTTGGTTA	AACCAGCGGT	4451
ACATTAATGC	AAGGGTTCAA	AACTACCGTG	CAGGCACTTT	CTAAATACTA	4401
CAATGTGACA	TTAATGCCGC	GCAGGAAGTA	TGGTAGCGTT	CAGCTCAGGA	4351
GTAAATCTTT	CAAGGGTCAG	TTACTTCAGC	AGTTCACACA	TACCGAAGCT	4301
GCAAATTAAC	ACATCATCGG	AACCTTAACT	AAGGAGCTGC	AATGCGACAG	4251
CGCAGAAATT	TTGGGAATGG	GATTTAACAG	AAACGCTGGC	ATGTTACGGC	4201
AATACGGTAA	GATTTCCGGT	TTGGTGGTAC	ACAGGTACAA	AACAAGTGCA	4151
AGGCTAACGT	ACAACAGGCG	AATTAAAGCA	CCAATTCAAA	ACCACTCAAT	4101
CGAAAGTTTA	TTAAAGCAAC	ACAGTAGAGG	TTCTGGTGGC	GCGGTACGAT	4051
GGCGATATCG	AAGTCAATCA	TAACCACTTC	ACCGAGAGTG	AATTAAAGGA	4001

6/68

### 7/68

## FIG. 1G.

			ልሞሞልሞር	АСАССТПАТТ АТТАТС	5101
AGTATTTTA	A TCTTGTAAAA AATTACGGAG AATACAATAA AGTATTTTTA	AATTACGGAG	TCTTGTAAAA	GCTTTACCCA	5051
TTCAGTACGG	AGTCATTTTA TTTTCGTATT ATTTACTGTG TGGGTTAAAG TTCAGTACGG	ATTTACTGTG	TTTTCGTATT	AGTCATTTTA	5001
CCTGCAATGA	A GCGGTCAGTA ATTGACAAGG TAGATTTCAT CCTGCAATGA	ATTGACAAGG	GCGGTCAGTA	ACGGGCGGTA	4951
ATCGCTGATA	A AACAGTGATG GCGCGACGGT GTGCGTTAAT ATCGCTGATA	GCGCGACGGT	AACAGTGATG	GTGTTTCTCA	4901
AAGGCAGGGC	A CCAGACCATT AAGTCGAATA GTGATTTCTG AAGGCAGGGC	AAGTCGAATA	CCAGACCATT	GAATTTGCAA	4851

# HIGH MOLECULAR WEIGHT FIG. 2A. AMINO ACID SEQUENCE OF PROTEIN I

					8/8	68									
KVRHLALKPL	IIRNSVDAII	DSNGQVFLIN	DKALAEIVNH	ISDIINPTIT	VSKDKSGNIV	IDLSGKEGGE	IVWGDIALID	DFDNVSINAE	LKKGTFVNIT	DTRGANLTIY	GTITSGNQKG	VNISMVLPKN	LTQPYNLNGI	ISVSGGGSVD	
EKGSEKPARM KVRHLALKPL	ATMQVDGNKT	NQISQLKGIL	ARNFTFEQTK	ISLLAGQKIT	RNQGKLSADS	DKVTLKTGAV	VSGKEKGGRA	AIVDAKEWLL	TLTNTTLESI	EINNDITTGD	KGSNQVITGQ	FEGTLNISGK VNISMVLPKN	DSRGSDSAGT	SLNYASFNGN	
S KRLNALVAVS ELARGCDHST	LQGMDVVHGT	NSAVFNRVTS	TLDISNENIK	EGVISVNGGS	GNINVRAATI	GGVISAQNQQ AKGGKLMITG	TSLEKGSTIN	SGHDLFIKDN	STPKRNKEKT	WSEGRSGGGV	ITAKQDIAFE	RTNKYAITNK	SESGEFNLTI	KAPIGINKYS	
KRLNALVAVS	SIPQSVLASG	EMVQFLQENN	IINTNGFTAS	VNLIGGKVKN EGVISVNGGS	VNLGDIFAKG	GGVISAQNQQ	GKNGIQLAKK	IAKTGGFVET	DEYTGSGNSA	INLSNGSLTL	ISLGAQGNIN	TGSGLQFTTK	TYWNLTSLNV	ERNARVNFDI	
MNKIYRLKFS	SAMLLSLGVT	NWKQFNIDQN	PNGITIGKDA	GLITVGKDGS	YSIAAPENEA	LSAKEGEAEI	TYLGGDERGE	GNINAQGSGD	TAGRSNTSED	ANQRIYVNSS	SGGWVDVHKN	FRFNNVSLNG	ESGYDKFKGR	SFNKDTTFNV	
<del>←</del> 1	51	101	151	201	251	301	351	101	151	501	551	501	551	01	

## FIG. 2B

						9	/68								
GFSIEKDLTL	RKAVTEIEGN	NIVNIAGNLT	IDNSKNLSIT	QKEGNLTISS	QDLNISGFNK	GHKVTLHSKV	SATSGEITTK	LAVSNISGNT	TVEVKATESL	DLTVGNGAEI	AGSINAANVT	VVNATNANGS	IDVKYIQPGI	NNTITVDTQN	•
RFKTSGSTKT	FEGGNITFGS	INSGNLTAGG	IAKGGARFKD	TEMQIGGDVS	TIKTKELKLT	QVKDSKISAD	NITSHKAVSI	SVTLTATEGA	GDIGGTISGG	NTVNVTANAG	SSHITSAKGQ VNLSAQDGSV	LNGAALGNHT	INTVLLKGVK	GVSAVRFIEP	IADNGR
YFNVSTGSSL	KGIVAKKNIT	KPLTIKKDVI	FDNKGNSNIS	DLNITNEGSD	DSDATNNANL	GTNAKKVTFN	IDAKNVTVNN	ILGGIESSSG	TESVTTSSQS	TGTIGGTISG	SSHITSAKGQ	TLVINAKDAE	NGLNIISKNG	DEEREALAKL	NSDGATVCVN
FTLLASSSNV QTPGVVINSK	QVEGTDGMIG	LIGSDFDNHQ	TNFTFNVGGL	ISGNITNKNG	IKAGVDGENS	LTIGNTNSAD	DSSDNNAGLT	NVEITAQTGS	TTLAGSTIKG	TTGEANVTSA	TSSGKLTTEA	KGSNINATSG	VNITGDLITI	RILEKVKDLS	VISEGRACFS
FTLLASSSNV	NATGGNITLL	VTINNNANVT	VESNANFKAI	TNSSSTYRTI	DKINITKQIT	AEITAKDGSD	ETSGSNNNTE	TGTTINATTG	VTVTANSGAL	TTQSNSKIKA	NATEGAATLT	LNTTGTLTTV	GSVIATTSSR	ASVDEVIEAK	EFATRPLSRI
751	801	851	901	951	1001	1051	1101	1151	1201	1251	1301	1351	1401	1451	.501

## HIGH MOLECULAR WEIGHT OF FIG. 3A. AMINO ACID SEQUENCE PROTEIN II (HMW2)

<del></del> 1	TAAATATACA	AGATAATAAA	AATAAATCAA	AGATAATAAA AATAAATCAA GATTTTTGTG	ATGACAAACA	
51	ACAATTACAA	CACCTTTTTT	GCAĞTCTATA	TGCAAATATT	TTAAAAAAT	
101	AGTATAAATC	CGCCATATAA	CGCCATATAA AATGGTATAA	TCTTTCATCT	TTCATCTTTA	
151	ATCTTTCATC	TTTCATCTTT	CATCTTTCAT	CTTTCATCTT	TCATCTTTCA	
201	TCTTTCATCT	TTCATCTTTC	ATCTTTCATC	TTTCATCTTT	CACATGAAAT	
251	GATGAACCGA	GGGAAGGGAG	GGAGGGGCAA	GAATGAAGAG	GGAGCTGAAC O	
301	GAACGCAAAT	GATAAAGTAA	TTTAATTGTT	CAACTAACCT	TAGGAGAAAA 00	
351	TATGAACAAG	ATATATCGTC	TCAAATTCAG	CAAACGCCTG	AATGCTTTGG	
401	TTGCTGTGTC	TGAATTGGCA	CGGGGTTGTG	ACCATTCCAC AGAAAAAGGC	AGAAAAAGGC	
451	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	CACTTAGCGT	TAAAGCCACT	
501	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	ATCTATTCCA	CAATCTGTTT	
551	TAGCAAGCGG	CTTACAAGGA	ATGGATGTAG	TACACGGCAC	AGCCACTATG	
601	CAAGTAGATG	GTAATAAAAC	GTAATAAAAC CATTATCCGC AACAGTGTTG	AACAGTGTTG	ACGCTATCAT	
651	TAATTGGAAA	CAATTTAACA	TCGACCAAAA	TGAAATGGTG	CAGTTTTTAC	
701	AAGAAAACAA	CAACTCCGCC	CAACTCCGCC GTATTCAACC GTGTTACATC	GTGTTACATC	TAACCAAATC	

## FIG. 3B

						11	/68	•							
TTTTAATCAA	ACTAATGGCT	GGCGCGTAAT	TTGTGAATCA	ATTGGTGGCA	CATTTCTTTA	CAACCATTAC	GGCGATATTT	TCGAAACCAA	GCAATATTGT	ATTTCCGCTC	CGATAAAGTC	AAGGGGGAGA	GGCATTCAAT	TGTATCAGGC	CGTTAATTGA
AGATTCTAAC GGACAAGTCT TTTTAATCAA	ATCACAATAG GTAAAGACGC AATTATTAAC	ATTTCTAACG AAAACATCAA GGCGCGTAAT	TTCACCTTCG AGCAAACCAA AGATAAAGCG CTCGCTGAAA TTGTGAATCA	TGTAAATCTT	ATTAGCGTAA ATGGTGGCAG	ATAATAAACC	GGTCAATCTG	CTGCCACTAT	GATAAAAGCG	TGGCGGTGTA	TGATTACAGG	TCAGGTAAAG AAGGGGGAGA	GGCGGTGACG AGCGCGGCGA AGGTAAAAAC GGCATTCAAT	CAACCATCAA	TATTGTGTG GGCGATATTG CGTTAATTGA
AGATTCTAAC	GTAAAGACGC	ATTTCTAACG	AGATAAAGCG	AAGACGGCAG	ATTAGCGTAA	AAAAAATCAC CATCAGCGAT	AAAATGAAGC	AATGTCCGTG	TGTAAGCAAA	AAGCGGAAAT	AGCTAAAGGC GGCAAGCTGA	TATCGACCTT	AGCGCGGCGA	GAAAAAGGCT	TATTGTGTGG
TCCCAATTAA AAGGGATTTT	ATCACAATAG	TACGCTAGAC	AGCAAACCAA	ACTGTCGGTA	CGAGGGTGTG	AAAAAATCAC	<b>೧೯೮೮</b>	CGGTAACATT	CTGCTGATTC	AAAGAGGGTG	AGCTAAAGGG	CAGGTGCAGT	GGCGGTGACG	AACCTCTTTA GAAAAAGGCT	GCGGACGCGC
TCCCAATTAA	CCCAAATGGT	TTACGGCTTC	TTCACCTTCG	CGGTTTAATT	AAGTGAAAA	CTCGCAGGGC	TTACAGCATT	TTGCCAAAGG	GGTAAACTTT	TCTTTCCGCC	AAAATCAGCA	ACATTAAAAA	AACTTACCTT	TAGCAAAGAA	AAAGAAAAAG
751	801	851	901	951	1001	1051	1101	1151	1201	1251	1301	1351	1401	1451	1501

## FIG. 3C.

	ATACATTTCA	CTCTTAATCT AGAGACAGGC GCAAATTTTA CCTTTAATTAA ATACATTTCA	GCAAATTTTA	AGAGACAGGC	CTCTTAATCT	2351
	AACGTCAGTG	TTCGCACTGG	CCAGCCATGA	TATTGGCAAA	GAACACCTCG	2301
	CTACGAGAAA	ATTAACCAAA	GAATATAACA	ACATATCTGG	GGCACAATTA	2251
	CAATCTTAGT	ATTTAACCCA	TCAGTGAATA	TATCATTTCA	AAGGTCTGAA	2201
	GGAACGGGTA	ATCTTTAAAC	CTAACAACGT	GATTTCAGGG	AGAGGGAAAA	2151
	CCATTACAGG	GGCACTGTAA	TGTCGCCCAG	ATGCTAAAAT	GACGCGGCAA	2101
	CAAAGCACGC	GTGGAAATAA	GCTTTTGAAG	CGCTTCCGTA	ATATTACCGC	2051
	GGTTTTTAA	GCTTGATCAG	AAAATATTAC	GATGTTCATA	CGGATGGGTT	2001
	TTTATTCTGG	AATTTAACCA	TAAAGGCGGA	ATATTACTTC	ATTGATGGAG	1951
1	AGGCGTTCAG	AGCGTGGCGG	AGTAAAGGTC	AATTCTCCAT	ACTCCCACTT	1901
	ATCGGAAGCA	AGAAAACTTA CCGTTAATAG CTCAATCAAC	CCGTTAATAG	AGAAAACTTA	AACGGCATCA	1851
	CAATGAATAT	ACCAATACAA CTATTTCAAATTATCTGAAA AACGCCTGGA CAATGAATAT	FTATCTGAAA 1	CTATTTCAAA	ACCAATACAA	1801
	AACAACGCTA	AAGCGACCCT AAAAAAATA GCGAACTCAA AACAACGCTA	AAAAAAATA	AAGCGACCCT	CCGGTGAAGC	1751
	CCAACAGGCA	TGATGAATTC	CCGGTATAAA	CGCAATAATA	AGACCCCCTT	1701
	TTGAAGCCGA	GATGTAACAA	AGACCCTGAT	AGTGGTTGCT	AAAACAAAAG	1651
	TGCAATTGTT	TTGACAGCAA	TATTTATCCA	ATCGGGGCAT	TTGTGGAGAC	1601
	ACCGGTGGTT	TATCGCTAAA ACCGGTGGTT	AACGCTCAAG GTAGTGGTGA		CGGCAATATT	1551

12 /68

#### FIG.3D

						13	/6	8							
CAGGGGTGAA	CAATCTCAAA GAAGGAGCGA	AAGCAAACCT	GGGCTCTGT	GAGTTAAAAA	AAATTCCCAT	CCATAAATGC	TTTTATGACG	CATTCTGGGC	TTACGGGGAA	AATAACGCCC	CAGCTTGCTC	TTAAAGGCAA	AGAGATACCC	AATTAATATA	GTGATTTAAA
AGAAGCTCTG CAGGGGTGAA	CAATCTCAAA	ACATGAACAC AAGCAAACCT	GCCACTGGTG	CAGAGGGGCT	AACGGCGCTA ATTTTACCTT	TAAAATCAAC AAAGACTTAA	TCAGACAGAC GAAAGATGAT	ACAACATATC	AGCAGCAGCA	GCTAGAAGCC	TAAAACTTGG	AATGCAGATA	AGGAAAGACT	GCACTGCCGA AATTAATATA	ACCAATGATG GTGATTTAAA
AACACAGTAT	ACATGTCATT	CCAAACGAGA	CAATATCACA	ACCATTCTGG	AACGGCGCTA	TAAAATCAAC	TCAGACAGAC	AATTCAACCT	GGTAATGTCA CCCTTGGTGG ACAAAACTCA AGCAGCAGCA	CAAATGTTAC	GATAGAGTTA	AACTGGCGAA	CCACTTTTAA	ACCAATAATG	TGGCAATGTT
AGCAATAGCA AAGGCTTAAC AACACAGTAT	GTAAATGGCA	AAGTTAATTT CAAATTAAAA	TTACCAATTC GGTTTTTAGC	ATATATGCCA	TAATATCTCT	GTTCGCGGCG ATGACGCTTT	AATTTCAGCC	GGTACGCACG CAATGCCATC	CCCTTGGTGG	TATTACTATC GAGAAAGCAG	CTAATCAGCA AAACATAAGG	GTTAATGGGA GTTTAAGTTT	TCAGAAAGCG	CGGCAATTTT	ACACAAGGAG TGGTAAAACT TGGCAATGTT
AGCAATAGCA	TTTTAACGGC	AAGTTAATTT	TTACCAATTC	TTTTTTTGAT	TGAGTGAAAT	GTTCGCGGCG	AACCAATTCA	GGTACGCACG	GGTAATGTCA	TATTACTATC	CTAATCAGCA	GTTAATGGGA	TCTCACTATT	TAAATATCAC CGGCAATTTT	ACACAAGGAG
2401	2451	2501	2551	2601	2651	2701	2751	2801	2851	2901	2951	3001	3051	3101	3151

#### FIG. 3E.

						1	4/6	68							
GGCGGAGATA	TAATGATGCT	ACCTCACGAT	AAAAAGGGTA	CAACCTAACT	TTTCAGGTTT	ACTATTGGCA	GGTGCCGAAG CCAAAACAGT AACTTTTAAC	TGACACTAAA	AGCAATAGCG	AGTAAACAAA	AAAAGGTTAC	GCAAGTATTA	CACGGTAAGT	AAATTGAAGC	ATTGGCGGTA
CACGCTAAAC GCAACCAAAG AAGCATCATC	CAGACAGTAA	TATCTCGCAA AAAGAAGGCA	TCACCAAACA GATAACAATC	TCAGATGCGA CAAGTAATGC	GACCTAAGTA	TAGAGATTTA	CCAAAACAGT	CTCTGCTGAC GGTCACAATG	CGGACGTGAA	ATTACTGCAA AAAATGTAGA AGTAAACAAA	CTCTCAAAAC AGTAAATATC ACCGCGTCGG AAAAGGTTAC	AAATGGCAAA GCAAGTATTA	TTTCCGGTAA CACGGTAAGT	TCCGGCTCAA AAATTGAAGC	GAGGCTAATG TAACAAGTGC AACAGGTACA ATTGGCGGTA
GCAACCAAAG	TTAAATATTA	TATCTCGCAA	TCACCAAACA	TCAGATGCGA	AAGAATTGAA ATTGACAGAA GACCTAAGTA	CCAAAGATGG	GGTGCCGAAG	CTCTGCTGAC	GCAGCAATGG	ATTACTGCAA	AGTAAATATC	TTAACGCAAC	AGCGGTACGA	CTGGTGATTT AACCACTAAA	TAACAAGTGC
CACGCTAAAC	AAAAGGAAGC	TTGGCGGCAA	AAAATTAATA	GGACTCTAGT	AAGAATTGAA	GAGATTACAG	CGGTAACAGC	ATTCAAAAAT	AAAACATCTA	CGGCTTAACT	CTCTCAAAAC	GGCTCGACCA	AGGTGATATC	CTGGTGATTT	GAGGCTAATG
CATTACCACT	TAATCAACAA	GAAATCCAAA	TTCTTCCGAT	TTGATGGAGA	ATTAAAACCA	CAATAAAGCA	ACAGTAATGA	AATGTTAAAG	TAGCAAAGTG	ACAACGATAC	GATATTACTT	CACCACAGCA	CAACCAAAAC	GTTAGCGCGA	GAAATCGGGT
3201	3251	3301	3351	3401	3451	3501	3551	3601	3651	3701	3751	3801	3851	3901	3951

#### FIG. 3F

						1	5/	68								
CGATTTAACA	CAACCTTAAC	ATCACTTCAA	CGCAGGAAGC	TAACCACCGT	ATTAACGCAA	AGAAGTGAAT	CCTCAAGCAG.	AAATGGGTTA AATATCATTT ®	AATTGAGGTG	TTGAAGCGAA	TGATGAAGAA AGAGAAACAT	TTGTTGAGCC AAATAATACA	CAAGTCAAGT	GGCGCACGAG	AATTGACAAG	TATTTACTGT
CAATTTCCGG TAATACGGTA AATGTTACGG CAAACGCTGG CGATTTAACA	GTTGGGAATG GCGCAGAAAT TAATGCGACA GAAGGAGCTG CAACCTTAAC	CTACTGAAGC CGGTTCTAGC	ATGGTAGCAT	ATTAAATACT ACAGGCACCT	GGCAGGCTCG GATATTAAAG CAACCAGCGG CACCTTGGTT	GTGATAGTAC AGAAGTGAAT	ACTGCGGCAA	TAAACACAGT AAATGGGTTA	GTGCGCTTAA GAGGCAAGGA AATTGAGGTG	AAATATATCC AGCCAGGTGT AGCAAGTGTA GAAGAAGTAA TTGAAGCGAA	TGATGAAGAA	TTGTTGAGCC	ATACACAAAA TGAATTTACA ACCAGACCGT CAAGTCAAGT	AAGTGGTAAT	CGTAGTCAGT	GTAGATTTCA TCCTGCAATG AAGTCATTTT ATTTTCGTAT TATTTACTGT
AATGTTACGG	TAATGCGACA	CTACTGAAGC	TTGGCTCAGA		CAACCAGCGG	GATGCATCAG	TGGTAGTGTG	TAAACACAGT	GTGCGCTTAA	AGCAAGTGTA	AAGATTTATC	GCTGTACGTT	TGAATTTACA	GATAATTTCT GAAGGTAAGG CGTGTTTCTC AAGTGGTAAT	GATGGACAGC	AAGTCATTTT
TAATACGGTA	GCGCAGAAAT	AATACCTTGA	CTAAGGGTCA GGTAGACCTC	CTAATGTGAC	GATATTAAAG		CAAGCGGCTC	ACTGGGGATT	TAGAAACACT	AGCCAGGTGT	ACGCGTCCTT GAAAAAGTAA AAGATTTATC	TGGTGTAAGT	ATACACAAAA	GAAGGTAAGG	TGTTGCTGAC	TCCTGCAATG
CAATTTCCGG	GTTGGGAATG	CGCAACAGGG	CTAAGGGTCA	ATTAATGCTG	GGCAGGCTCG	AAGATGCTAA GCTAAATGGT	GCAGTCAACG CAAGCGGCTC	TGTGAATATC	CGAAAGATGG	AAATATATCC	ACGCGTCCTT	TAGCTAAACT	ATTACAGTCA	GATAATTTCT	TATGTACCAA	GTAGATTTCA
4001	1051	1101	1151	1201	1251	1301	1351	1401	1451	1501	1551	1601	1651	1701	1751	1801

#### 16/68

# FIG. 3G.

GTGGGTTAAA GTTCAGTACG GGCTTTACCC ATCTTGTAAA AAATTACGGA GAATACAATA AAGTATTTTT AACAGGTTAT TATTATG 4851 4901

BNSDOCID: <WO\_\_\_9421290A1\_J\_

# FIG.4A. AMINO ACID SEQUENCE OF HIGH MOLECULAR WEIGHT PROTEIN

ELARGCDHST EKGSEKPARM KVRHLALKPL

KRLNALVAVS

MNKIYRLKFS

17/68 DSNGQVFLIN DKALAEIVNH ISDIINPTIT GNINVRAATI RNQGKLSADS VSKDKSGNIV IDLSGKEGGE IVWGDIALID SGHDLFIKDN AIVDAKEWLL DFDNVSINAE KGGNLTIYSG ASVAFEGGNN KARDAANAKI VAQGTVTITG TQYRSSAGVN IIRNSVDAII YLKNAWTMNI NITINOTTRK SKPLPIRFLA NITATGGGSV GKNGIQLAKK TSLEKGSTIN VSGKEKGGRA SDPKKNSELK TTLTNTTISN SIPQSVLASG LQGMDVVHGT ATMQVDGNKT NQISQLKGIL TLDISNENIK ARNFTFEQTK EGVISVNGGS ISLLAGQKIT DKVTLKTGAV ILHSKGQRGG GVQIDGDITS IISSVNNLTH NLSGTINISG YISSNSKGLT GGVISAQNQQ AKGGKLMITG ETGANFTFIK KLKPNENMNT EMVQFLQENN NSAVFNRVTS VNLIGGKVKN SINIGSNSHL LDQGFLNITA SHWNVSALNL IINTNGFTAS VNLGDIFAKG DEFPTGTGEA SLNGTGKGLN NLKEGAKVNF IAKTGGFVET NWKQFNIDQN GLITVGKDGS TYLGGDERGE SAMLLSLGVT PNGITIGKDA DPLRNNTGIN TASRKLTVNS GWVDVHKNIT NTSYWQTSHD FNGVNGNMSF YSIAAPENEA GNINAQGSGD EGKDFRANNV LSAKEGEAEI 51 201 301 401 451 551 101 151 251 351 501 601 651 701

#### FIG. 4B

					•	18/	68							
KINITH INTINIA	ONSSSSITGN	TGENADIKGN	INITQGVVKL GNVTNDGDLN	ISQKEGNLTI	LTEDLSISGF	SADGHNVTLN	VNITASEKVT	TTKSGSKIEA	NATEGAATLT	LNTTGTLTTV	GSVTAATSSS	ASVEEVIEAK	EFTTRPSSQV	
NCHIDCHIDE	ILGGNVTLGG	SLLVNGSLSL	INITQGVVKL	NDAEIQIGGN	DSSSDATSNA NLTIKTKELK	TFNNVKDSKI	VNKDITSLKT	TVSVSATVDL	DLTVGNGAEI	AGSINAANVT	EVNAVNASGS	IEVKYIQPGV	NNTITVNTQN	
NT SNCANETI.		NIRDRVIKLG	GNFTNNGTAE	KGSLNITDSN	DSSSDATSNA	GNSGAEAKTV	GLTITAKNVE	GDISGTISGN	NTVNVTANAG	VDLLAQNGSI	LVINAKDAK LNGDASGDST	RNTVRLRGKE	GVSAVRFVEP	VADDGQP
RGAEL KMSET	KDDFYDGYAR	LEANNAPNQQ	GKTRDTLNIT	SIIGGDIINK	ITIKKGIDGE	RDLTIGNSND	GRESNSDNDT	NGKASITTKT	TGTIGGTISG	GSSITSTKGQ VDLLAQNGSI	TLVINAKDAK	NGLNIISKDG RNTVRLRGKE	DEERETLAKL	SGNGARVCTN VADDGQP
FFDIYANHSG	TNSNFSLRQT	ITIEKAANVT	LTISESATFK	ITTHAKRNQR	SSDKINITKQ	NKAEITAKDG	SKVKTSSSNG	TTAGSTINAT	KSGEANVTSA	ATGNTLTTEA	AGSDIKATSG	VNITGDLNTV	RVLEKVKDLS	IISEGKACFS
751	801	851	901	951	1001	1051	1101	1151	1201	1251	1301	1351	1401	1451

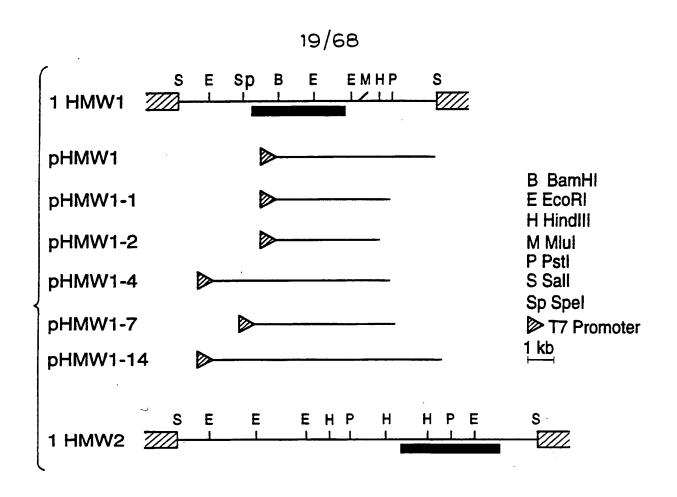
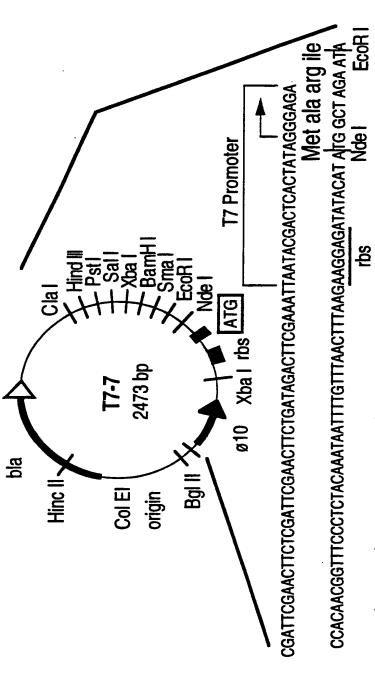


FIG.5A.





arg ala arg gly ser ser arg val asp leu gln pro lys leu ile ile asp .. cec ecc dee eea tob tot ada etc eab die cae ccc aae cifi atc atc dat ... SallPstl Smal BamHI Xbal

F16.5B.

shaded boxes indicate the locations of the structural genes. In the recombinant phage, transcription proceeds from left to right for the HMW1 gene and from right to left for the HMW2 gene. The methods used for construction of the plasmids shown are (A) Partial restriction maps of representative HMW1 and HMW2 recombinant phage and of HMW1 plasmid subclones. The described in the text. (B) Restriction map of the T7 expression vector pT7-7. This vector contains the T7 RNA polymerase promoter 410, a ribosome - binding site (rbs), and the translational start site for the T7 gene 10 protein upstream from a

multiple cloning site (37).

#### FIG. 6A

ATGACAAACA	TTAAAAATA	TCATCTTTCA	CATCTTTCAT	ACATGAAATG	CTGAACG	AGGAGAAAAT ~	ATGCTTTGGT ®	AAAGGCA	AAAGCCACTT	AATCTGTTTT	ACTATGC	CGCTATCATT	AGTTTTTACA	CAAATCT	
	TGCAAATATT TT?	CTTTCATCTT TC2	TTTCATCTTT CA1	TTCATCTTTC ACA	GAGGGCCAAG AATGAAGAGG GAGCTGAACG	AACTAACCTT AGG	AAACGCCTGA ATG	CCATTCCACA GAAAAAGGCA	ACTTAGCGTT AAA	TCTATTCCAC AAI	ACACGGCACA GCCACTATGC	ACAGTGTTGA CGC	GAAATGGTGC AGI	TGTTACATCT AACCAAATCT	
GTACAAACCC ACAATAAAAT	GCAGTCTATA 1	ATGGTATAAT C	ATCTTTCATC 1	TCTTTCATCT 1	GAGGGGCAAG A	TTAATTGTTC A	CAAATTCAGC A		AAAGTGCGTC A	AGGTGTAACA 1	TGGATGTAGT A	ATTATCCGCA A	CGACCAAAAT C	TATTCAACCG I	
ACAGCGTTCT CTTAATACTA	CACCTTTTT	CC GCCATATAAA	TTCATCTTTC	TCATCTTTCA	ATGAACCGAG GGAAGGGAGG	ATAAAGTAAT	TATATCGTCT	GAATTGGCAC GGGGTTGTGA	TGCTCGCATG	TACTATCTTT	TTACAAGGAA	AAGTAGATGG TAATAAAACC	AATTGGAAAC AATTTAACAT	AGAAAACAAC AACTCCGCCG	
ACAGCGTTCT	ACAATTACAA	GTATAAATCC	TCTTTCATCT	CTTTCATCTT	ATGAACCGAG	AACGCAAATG	ATGAACAAGA	TGCTGTGTCT	GCGAAAAACC	TCCGCTATGT	AGCAAGCGGC	AAGTAGATGG	AATTGGAAAC	AGAAAACAAC	
$\vdash$	51	101	151	201	251	301	351	401	451	501	551	601	651	701	751

#### FIG. 6B

	GITAATTIGAC CCGGTGGTTT GCAATTGTTG		TAGTGGTGAT ATTTATTCAT	ACGCTCAAGG TCGGGGCATG	GGCAATATTA TGTGGAGACG	1551 1601
	GTTAATTGAC	GCGATATTGC	ATTGTGTGGG	CGGACGCGCT	AAGAAAAAGG	1501
	GTATCAGGCA	ACCTCTTTAG AAAAAGGCTC AACCATCAAT	AAAAAGGCTC	ACCTCTTTAG	AGCAAAGAAA	1451
	GCATTCAATT	GGTAAAAACG GCATTCAATT	GCGCGGCGAA	GCGGTGACGA	ACTTACCTTG	1401
	AGGGGGAGAA	CAGGTAAAGA	ATCGACCTTT	AGGTGCAGTT	CATTAAAAAC	1351
	GATAAAGTCA	GATTACAGGC GATAAAGTCA	GCTAAAGGCG GCAAGCTGAT	GCTAAAGGCG	AAATCAGCAA	1301
8	TTTCCGCTCA	GGCGGTGTAA	AGCGGAAATT	AAGAGGGTGA	CTTTCCGCCA	1251
2/6	CGAAACCAAG	TGCCACTATT	ATGTCCGTGC	GGTAACATTA	TGCCAAAGGC	1151
2	GCGATATT'I'	GTCAATCTGG	CCGCCCTGA AAATGAAGCG	CCGCGCCTGA	TACAGCATTG	1101
	AACCATTACT	TAATAAACCC	ATCAGCGATA	AAAAATCACC	TCGCAGGGCA	1051
	ATTTCTTTAC	TGGTGGCAGC	TTAGCGTAAA	GAGGGTGTGA	AGTGAAAAAC	1001
	TTGGTGGCAA	GTAAATCTTA	AGACGGCAGT	CTGTCGGTAA	GGTTTAATTA	951
	TGTGAATCAC	TCGCTGAAAT	GATAAAGCGC	GCAAACCAAA	TCACCTTCGA	901
	GCGCGTAATT	TTTCTAACGA AAACATCAAG	TTTCTAACGA	ACGCTAGACA	TACGGCTTCT	851
	CTAATGGCTT	TCACAATAGG TAAAGACGCA ATTATTAACA CTAATGGCTT	TAAAGACGCA	TCACAATAGG	CCAAATGGTA	801

#### FIG. 6C.

CAAGAGTCAA	GAACGAAATG	CTTTAATGTT	AAGACACTAC	TCATTCAACA	451
AAACGGTATA	CTTATAATTT	CTTACCCAGC	TGCAGGCACA	GAAGCGATAG	401
GACTCCAGAG	CCTCACTATT	CAAAGGACGC	ATGATAAATT	GAAAGTGGAT	351
ATTTAACCTC	ACTTACTGGA	CAAAGGACGC	ATGATAAATT	GAAAGTGGAT	301
ACCTAAAAAT	CAATGGTTTT	GTGAACATCT	TTCAGGGAAA	CTTTAAATAT	251
TTTGAAGGGA	CACAAATAAA	AATACGCTAT	AGAACCAATA	CACCACTAAA	201
GACTGCAATT	ACTGGCAGCG	TCTAAACGGC	ATAATGTCTC	TTTAGATTTA	151
TCAAAAAGGT	CCTCAGGCAA	GGGACTATTA	TACAGGTCAA	ACCAAGTCAT	101
AAAGGAAGCA	CGCCTTTGAG	AACAAGATAT	ATTACAGCTA	TAACATAAAC	051
GGGCGCAAGG	ATCTCACTCG	TCATAAAAAT	GGGTTGATGT	TCAGGCGGCT	001
AACAATTTAC	GTGCAAACTT	GATACCAGAG	CACCGGTGAT	ACGATATTAC	.951
GAGATTAACA	TGGCGGCGTT	GTCGGAGCGG	TGGAGTGAGG	CTTAACTCTT	901
CCAATGGCAG	ATTAATTTAT	CAATAGCTCC	GCATCTATGT	GCTAATCAAC	851
TAACATCACT	GTACCTTTGT	CTAAAAAAAG	TGAGAGTATA	ACACAACTCT	.801
ACATTAACAA	AGAAAAGACA	AACGAAACAA	AGCACCCCAA	GAATAGTGCC	.751
CGGGATCCGG	GATGAATACA	TTCAGAAGAC	GCAGCAATAC	ACAGCAGGAC	.701
TAATGCAGAA	ATGTATCTAT	GACCCGGATA	GTGGTTGTTA	ACGCCAAAGA	.651

#### FIG. 6D

2501	CTTTGACATC	AAGGCACCAA	TAGGGATAAA	CTTTGACATC AAGGCACCAA TAGGGATAAA TAAGTATTCT AGTTTGAATT	AGTTTGAATT
2551	ACGCATCATT	TAATGGAAAC	ATTTCAGTTT	CGGGAGGGGG	GAGTGTTGAT
2601	TTCACACTTC	TCGCCTCATC	CTCTAACGTC	CTCTAACGTC CAAACCCCCG	GTGTAGTTAT
2651	AAATTCTAAA	TACTTTAATG	TTTCAACAGG	TTTCAACAGG GTCAAGTTTA AGATTTAAAA	AGATTTAAAA
2701	CTTCAGGCTC	AACAAAAACT	GGCTTCTCAA	TAGAGAAAGA	TTTAACTTTA
2751	AATGCCACCG	AATGCCACCG GAGGCAACAT	AACACTTTTG	AACACTTTTG CAAGTTGAAG	GCACCGATGG
2801	AATGATTGGT	AATGATTGGT AAAGGCATTG	TAGCCAAAAA	TAGCCAAAAA AAACATAACC	TTTGAAGGAG N
2851	GTAAGATGAG	GTTTGGCTCC	AGGAAAGCCG TAACAGAAAT	TAACAGAAAT	CGAAGGCAAT ®
2901	GTTACTATCA	GTTACTATCA ATAACAACGC	TAACGTCACT	TAACGTCACT CTTATCGGTT	CGGATTTTGA
2951	CAACCATCAA	CAACCATCAA AAACCTTTAA	CTATTAAAAA	CTATTAAAAA AGATGTCATC	ATTAATAGCG
3001	GCAACCTTAC	CGCTGGAGGC	AATATTGTCA	ATATAGCCGG AAATCTTACC	AAATCTTACC
3051	GTTGAAAGTA	GTTGAAAGTA ACGCTAATTT	CAAAGCTATC	CAAAGCTATC ACAAATTTTCA CTTTTAATGT	CTTTTAATGT
3101	AGGCGGCTTG	TTTGACAACA	AAGGCAATTC	AAATATTTCC	ATTGCCAAAG
3151	GAGGGGCTCG	CTTTAAAGAC	ATTGATAATT	CCAAGAATTT AAGCATCACC	AAGCATCACC
3201	ACCAACTCCA	ACCAACTCCA GCTCCACTTA	CCGCACTATT	ATAAGCGGCA ATATAACCAA	ATATAACCAA
3251	TAAAAACGGT	GATTTAAATA	TTACGAACGA	TTACGAACGA AGGTAGTGAT ACTGAAATGC	ACTGAAATGC

#### FIG. 6E

							1								
_				۰		25	/68	_	۰						
GATTTCTTCT	GTGTTGATGG	ACCATTAAAA	TTTCAATAAA	GTAACACCAA	CAGGTTAAAG	CAGCAAAGTG	ACAATAATGC	AATATTACTT	TACCACTAAA	TAACCGCTCA	TCTGTAACAC	GGGCAACACC	CAGGCTCTAC	GGCGATATCG	CGAAAGTTTA
AAATTGGCGG CGATGTCTCG CAAAAGAAG GTAATCTCAC GATTTCTTCT	ATCAAGGCAG	TGCCAATCTA ACCATTAAAA	ATATTTCAGG	TTAACTATTG	AACCTTTAAC CAGGTTAAAG	TGACACTACA	GTAGTAATAA CAACACTGAA GATAGCAGTG ACAATAATGC	ATCGATGCAA AAAATGTAAC AGTAAACAAC AATATTACTT	TCTGCGACAA GTGGAGAAAT	CCATTAACGC AACCACTGGT AACGTGGAGA	CAGCTCTGGC	GCAATATTTC GGGCAACACC	CTGCAAATAG CGGTGCATTA ACCACTTTGG CAGGCTCTAC	TAACCACTTC AAGTCAATCA GGCGATATCG	TTCTGGTGGC ACAGTAGAGG TTAAAGCAAC CGAAAGTTTA
CAAAAAGAAG	ATATTACCAA ACAGATAACA	GATTCAGACG CGACAAACAA	GAAATTAACG CAAGACCTAA	TGGTAGTGAT	GGTACTAATG CCAAAAAAGT	CTCTGCTGAC GGTCACAAGG	CAACACTGAA	AAAATGTAAC	TCTGCGACAA	AACCACTGGT	ATCCTAGGTG GAATTGAGTC	CTTGCTGTAA	CGGTGCATTA	TAACCACTTC	ACAGTAGAGG
CGATGTCTCG	ATATTACCAA	GATTCAGACG	GAAATTAACG	CAGCTAAAGA	GGTACTAATG	CTCTGCTGAC	GTAGTAATAA	ATCGATGCAA	AGTGAGCATC	CCATTAACGC	ATCCTAGGTG	CGAGGGCGCT	CTGCAAATAG	ACCGAGAGTG	TTCTGGTGGC
AAATTGGCGG	GACAAAATCA	GGAGAATTCC	CCAAAGAATT	GCAGAGATTA	TAGTGCTGAT	ATTCAAAAAT	GAAACATCCG	CGGCTTAACT	CTCACAAAGC	ACAGGTACAA	AACAGGTAGT	TTACTGCAAC	GTTACTGTTA	AATTAAAGGA	GCGGTACGAT
3301	3351	3401	3451	3501	3551	3601	3651	3701	3751	3801	3851	3901	3951	4001	4051

### FIG. 6F.

4101	ACCACTCAAT	CCAATTCAAA	AATTAAAGCA	ACCACTCAAT CCAATTCAAA AATTAAAGCA ACAACAGGCG AGGCTAACGT	AGGCTAACGT
4151	AACAAGTGCA	AACAAGTGCA ACAGGTACAA		TTGGTGGTAC GATTTCCGGT	AATACGGTAA
4201	ATGTTACGGC	AAACGCTGGC	GATTTAACAG	TTGGGAATGG	CGCAGAAATT
4251	AATGCGACAG	AAGGAGCTGC	AACCTTAACT	ACATCATCGG	GCAAATTAAC
4301	TACCGAAGCT	AGTTCACACA	TTACTTCAGC	TTACTTCAGC CAAGGGTCAG GTAAATCTTT	GTAAATCTTT
4351	CAGCTCAGGA	TGGTAGCGTT	GCAGGAAGTA	TTAATGCCGC	CAATGTGACA
4401	CTAAATACTA	CAGGCACTTT	AACTACCGTG	AAGGGTTCAA ACATTAATGC	
4451	AACCAGCGGT	ACCTTGGTTA	TTAACGCAAA	TTAACGCAAA AGACGCTGAG	CTAAATGGCG 0
4501	CAGCATTGGG	TAACCACACA	GTGGTAAATG	GTGGTAAATG CAACCAACGC AAATGGCTCC	
4551	GGCAGCGTAA	TCGCGACAAC		CTCAAGCAGA GTGAACATCA CTGGGGATTT	CTGGGGATTT
4601	AATCACAATA	AATGGATTAA	ATATCATTTC	ATATCATTTC AAAAAACGGT	ATAAACACCG
4651	TACTGTTAAA	AGGCGTTAAA	ATTGATGTGA	ATTGATGTGA AATACATTCA ACCGGGTATA	ACCGGGTATA
4701	GCAAGCGTAG	ATGAAGTAAT	TGAAGCGAAA	TGAAGCGAAA CGCATCCTTG AGAAGGTAAA	AGAAGGTAAA
4751	AGATTTATCT	GATGAAGAAA GAGAAGCGTT	GAGAAGCGTT	AGCTAAACTT GGCGTAAGTG	GGCGTAAGTG
4801	CTGTACGTTT	TATTGAGCCA	AATAATACAA	TTACAGTCGA	TACACAAAAT
4851	GAATTTGCAA	CCAGACCATT	AAGTCGAATA	GTGATTTCTG AAGGCAGGGC	AAGGCAGGGC
4901	GTGTTTCTCA	AACAGTGATG	GCGCGACGGT	AACAGTGATG GCGCGACGGT GTGCGTTAAT ATCGCTGATA	ATCGCTGATA

### FIG. 6G.

4951	ACGGCCGGTA	GCGGTCAGTA	GCGGTCAGTA ATTGACAAGG	TAGATTTCAT	CCTGCAATGA
5001.	AGTCATTTTA	TTTTCGTATT	ATTTACTGTG	TGGGTTAAAG	TTCAGTACGG
5051	GCTTTACCCA	TCTTGTAAAA	AATTACGGAG	AATACAATAA	AGTATTTTA
5101	ACAGGTTATT	ATTATGAAAA	ATTATGAAAA ATATAAAAAG	CAGATTAAAA CTCAGTGCAA	CTCAGTGCAA
5151	TATCAGTATT	GCTTGGCCTG	GCTTCTTCAT	CATTGTATGC	AGAAGAAGCG
5201	TTTTTAGTAA	AAGGCTTTCA	GTTATCTGGT	GCACTTGAAA	CTTTAAGTGA
5251	AGACGCCCAA	CTGTCTGTAG	CTGTCTGTAG CAAAATCTTT	ATCTAAATAC	ATCTAAATAC CAAGGCTCGC &
5301	AAACTTTAAC	AAACCTAAAA	ACAGCACAGC	TTGAATTACA	TIGAATTACA GGCIGIGCIA $^{\circ}_{\mathfrak{D}}$
5351	GATAAGATTG	AGCCAAATAA	GTTTGATGTG	ATATTGCCAC AACAAACCAT	AACAAACCAT
5401	TACGGATGGC	AATATTATGT	TTGAGCTAGT	CTCGAAATCA GCCGCAGAAA	GCCGCAGAAA
5451	GCCAAGTTTT	TTATAAGGCG	AGCCAGGGTT	ATAGTGAAGA	AAATATCGCT
5501	CGTAGCCTGC	CATCTTTGAA	CATCTTTGAA ACAAGGAAAA	GTGTATGAAG ATGGTCGTCA	ATGGTCGTCA
5551	GTGGTTCGAT	TTGCGTGAAT	TCAATATGGC	TCAATATGGC AAAAGAAAAT	CCACTTAAAG
5601	TCACTCGCGT	GCATTACGAG	TTAAACCCTA	TTAAACCCTA AAAACAAAAC	CTCTGATTTG
5651	GTAGTTGCAG	GTTTTTCGCC	TTTTGGCAAA	TTTTGGCAAA ACGCGTAGCT	TTGTTTCCTA
5701	TGATAATTTC	GGCGCAAGGG		AGTTTAACTA TCAACGTGTA	AGTCTAGGTT

#### FIG. 6H

CTGCGGGTTT	ACGGTATCCT	AGATATGCAC	CTTACGGCGA	AATGCTAAAA	5551
TAATAGCGAA	AGTTCCGTTA	GATGCAGGTC	TGCGTTTTAT	TCAGCCCTTA	5501
CGCTTTCAAA	AAAATACACC	TAAGTATGCC	CGTAATGAAT	TCTTGTATGG	5451
GTGAGCGCGG	GGTGCAAGTG	TAAATACGGC	TCAGAGGCTT	ACTTATGGCG	5401
TGTAACAGGT	ATTTATTCTC	AGTAGCATAG	ACAAGATATA	AGTTTACTCT	5351
TTATCGGGTC	TAGCAGTCAA	GTTGGCATTT	TTTGCTCAAG	GAGTCAAGAG	5301
GTTTAGGGTT	AGCACAGCCA	CTATCACATT	TTAATCGCAG	GGCGAAACAT	5251
GGAGCGCATT	CTTTTGGAAT	TTACCAGGCT	CGCGAGTAAA	ATCATTATTA	5201
GATTTAACTC	CTTTAATATT	CTAAAACAAT	CAATTTACCC	TGGACATATC	5151
CAGGCATTGA	GGCGTAAGTG	TGCAGTATCA	AGAAAAATT	GGTGCAACGA	5101
AAACACCCTG	CATCCGAGTT	ATTAATCAAA	CTACCGCCAT	TAGGCTACAA	5051
AAAATTAATT	AGACCAGTTT	TTGGAATGGA	ACATTTAACC	TTATCTCCCG	5001
AATGGAGTTA	GCGAATCTGA	ATCTATCTCT	CAAAAGGTCA	CGTAAATTAT	5951
TGCGATTAAT	GCTTACCAAG	GATATCGACG	TGATTCTAAT	TGAGTTATGC	5901
TATACCAGCA	CTTAAGTCTT	AACACCAATC	TTTTATGATA	TACTTATCCG	5851
GCATAGGATA	TATGCGGTAG	ATCAAAATCT	TAAAAGCACC	TTGACCAATG	5801
TCTAAACGCA	ATGTATTAAA	GGACATGATG	CAATTTGACC	TTGTAAATGC	5751

#### FIG. 6I

6751 AC 6801 AT 6851 TT 6901 AA 6951 AG 7001 AC 7051 GA 7151 TA 7201 AA 7251 AT	ACCCTGAAAT ATATGCTTTA ATATATATCA AGCAAACCAA ACAATTTATA ACAATTTAAA ACCGCTTCAC TAAAAAGATTA AAAAAGATTA	ACCCTGAAAT TTAATCAACT ATATGCTTTA CCCGCCAATT TTATATATCA AACAAACTAA AACCAAGCAA ACCAAGCAAA ACCAATTTATA TGATAAACTA GATTTAATAA TATGACAAAA ACCGCTTCAC TTGTAGAATC TAAACAACCA CCCAAACCCA AAAAAGATTA TGAGCTTGCT AAAAAGATTA TGAGCTTGCT	TTAATCAACT GGTAAGCGTT CCGCCTACCA GTTTATAACT CCCGCCAATT TACAGTCTAT ACGCAACCCT GTTTTCATCC AACAAACTAA GCAAACCAAG CAAACCAAGC ACCAAGCAAA CCAAGCAAAC AAGCAAACA GCAAACCAAG AAACCAAGC AAGCAAAAA TGATAAACTA AAACATACTC CATACCATGG CAATACAAGG TATGAAAAA GAAAATTTAC AAAGTGTTCC ACAAAATACT CCCAAACCCA ACCTATTAC CAAACTTCC TGCAAATACT CCCAAACCCA ACCTATTACG CCTGGAACAA TGAGCTTGCT TGCCGCGAAT TAATGGCGAT TTTGGAAAAA ATTTTGGAGG CGTTCACGAT ATTGAATTTG ACGCACCTGC	GGTAAGCGTT CCGCCTACCA GTTTATAACT TACAGTCTAT ACGCAACCCT GTTTTCATCC GCAAACCAAG CAACCAAGC AAGCAAACCA CCAAGCAAAC CAAGCAAACCA CAAACCAAGC AAGCTAAAAA AAACATACTC CATACCATGG CAATACAAGG GAAAATTTAC AAAGTGTTCC ACAAAATACG AAACAACGAC CAAACTTCCC TGCAAATACT ACCTATTACG CCTGGAACAA CATGTCGCCA TGCCGCGAAT TAATGGCGAT TTTGGAAAAA CGTTCACGAT ATTGAATTTG ACGCACCTGC	ACCCTGAAAT TTAATCAACT GGTAAGCGTT CCGCCTACCA GTTTATAACT ATATGCTTTA CCCGCCAATT TACAGTCTAT ACGCAACCCT GTTTTCATCC TTATATATATCA AACAAACTAA GCAAACCAAG CAAACCAAGC AACCAAGCA AACCAAGCAA CCAAGCAAA CCAAGCAAAC AAGCAAACC AAGCAAACCA GAACCTAAAAA GAAATTTATA TGATAAACTA GAAACTTTAC CATACCATG CAATACAAG GAAATTTAC AAAGTGTTCC ACAAAATACG ACCGCTTCAC TTGTAGAAA GAAAATTTAC AAAGTGTTCC TGCAAATACT TAAACAACCA CCCAAACCCA ACCTATTACG CCTGGAACACA CCCAAACCCA ACCTATTACG CCTGGAAACAAAAAAAAAA
	ATGGACGCTA TCAGCTGGCA	Al'I'I'I'GGAGG TATCTACCCG	AI"I"I"IGGAGG CGT"ICACGAT ATTGAATTTG TATCTACCCG AAAAACTACT AATTCATTTT		ACGCACCTGC GCCACTCGTC
	AGCTGGCA	TATCTACCCG	AAAAACTACT		GCCACTCGTC
7351 TC	SCTAATGC	AATTACAACA	TCGCTAATGC AATTACAACA CTCTTTTCCG ACCCCGAATT		GGCAATTTCC

# FIG. 6J.

						3C	/68	3							
TGACGCTGAT	AATAAATATA	CAACTCT	ATGTCAATAT	GCTTCATTGT	TGCGTTTCAT	AAATTGC	TATATGCACT 0	TCCATTAAAC	ACCGCTACCT	GTACTGCTTG	TCAATG	ATGAGGGCGT	ATCAGTAGCA	CGAAACTTTC	TTACCACGAT
TGA(	AAT	AGAC	ATG	GCT	TGC	ZDDD	TAT	TCCZ	ACCO	GTAC	AACT	ATGA	ATCA	CGAA	TTAC
GATTAGCCTG CAACGCTGGT	CCATATTCTC	ATTTAGCAAC AGACAACTCT	CCCGAATCCA	ACAACTTTGT	GTACTGCATC	AAAAAACTCG	TCATGATGTA	ATGTTAAGCG	GGATGGCAAG	TGTGATGATG	GCACGCATTC AACTTCAATG	GGCTTAGGCC	GTTCTTTGAA	GTAAACAGTG	TGTTCTATAT GCCAAGCATT GGCATGGATA
	TTAACGCAGA	GGTGGCTTTC	TTTTTACTTA	CAGGGAATCA	CGTTTTATTG	GTGGTTTCCT	CAAATATCCT	TTTAGCAAAA AACAAGCACG	CCTCACGCAA	ACGGCAAACC	TCGATTTATC	CTATTTAGTC	TGTTTGACGA	TTTTTATCC	GCCAAGCATT
GAAGAAGGGG CATTAAAGAT	TCCCCCTACG	AGATTCCGAA	AATTCTGTAT	GCGTTATGGG	GCAGTCTTCA	TGGTTTTACA	GAATTGCCTG	TTTAGCAAAA	GCAAGCATAT	GGTAAAAAGG	TTCGGGACAT	ATTGCTGCTC GAGAAAATT	GGTCGAGAAG	GGAGAGACTG	TGTTCTATAT
GAAGAAGGGG	TTTTGCCTCT	ATATCAACCC	TCTATTGCTA	GAGTTTAGAT	GTTTTGCGTT	AAAAGAGCGG	TAATTTAGAT	GCAGTTATGA	GAACTTGTCC	TTACACCTTA	AACATTTTAA	ATTGCTGCTC	TGATAACATA	ATAATATAAT	CAACCCGCAG
7401	7451	7501	7551	7601	7651	7701	7751	7801	7851	7901	7951	8001	8051	8101	8151

#### FIG. 6K

AATTTTCAAA	AAGTGCGGTT	TTTAAAGTAA	ATAACGGTTT	TGAGTAAAAA	9001
CGGAAGCACT	TGAATGGAAG	AGAAAACAAA	ATACTGCTTA	ATTGGGCAAA	8951
ACCCTCGTCC	TTTACAGGCG	ACAAAAGCTT	ACAACGGCTT	ATCATAGAAA	8901
CCGTCGTTAC	GCCTTGAACT	CATCAAGAAC	AGCAGAAAAC	CTTTGCGTCT	8851
ATTGAATGTG	AGAAACATAT	CCGACACACG	TGGCTGATAG	ACTACCAGAA	8801
AACGCTTAGG	GGTCTGTTTA	TATTGATGAA	TACATGAACA	GGGGATGAAG	8751
ATGCAAAACG	TAGTTGGTGT	ACATTAGGTT	TGATATGGTT	ACGGCATAAT	8701
GGTAATACTA	GTTTCCTTTC	TACTAAATCC	TGCGATATGC	ATTGCGTGAT	8651
ATCTGGCAAT	TATCACGATT	CCACGCACCT	CTGCACATCC	GACGATGCCA	8601
CTATTTAGGT	TTATCGAAAG	GTCAAATGGT	ACACCCTTAT	CAGGCTTGAC	8551
GGACAATCAA	TTTCGCACTT	TACATTTTCA	AAAGTCAAAA	AGATAAAGCT	8501
AAGAAATCAG	CTAACATTGC	TGAATTTTTG	AATTAAACCC	ACCACAATGA	8451
TATTGCCGCT	TCAATATCGG	CCTGAAGTAG	CAGGGAAAAC	ATTATGTACT	8401
CAAAAAGTGG	ACTCGCCCCA	TACCATCTGC	CTACCTTATG	CAAAGATGCC	8351
TACGCTTACC	GAAACCCTTT	TTGTTTAGC	GCAGTGAAGA	GATTATGTGG	8301
CGTAGAAGAT	ATTATGTCAT	GAATTTATTG	TACGCATTCT	ATCCTGCCAC	8251
GCCTTGGGTC	TCAAGCTGTA	TTGCCCCTAT	AACACTCGGC	TTTTGTGAGC	8201

32 /68

# FIG. 6L

#### FIG. 7A

<del></del>	CGCCACTTCA		ATTTTGGATT GTTGAAATTC AACTAACCAA AAAGTGCGGT	AACTAACCAA	AAAGTGCGGT	
51	TAAAATCTGT	GGAGAAAATA	GGTTGTAGTG	GGTTGTAGTG AAGAACGAGG	TAATTGTTCA	
101	AAAGGATAAA	GCTCTCTTAA	TTGGGCATTG	GTTGGCGTTT	CTTTTTCGGT	
151	TAATAGTAAA	TTATATTCTG	GACGACTATG	CAATCCACCA	ACAACTTTAC	
201	CGTTGGTTTT	AAGCGTTAAT	GTAAGTTCTT	GCTCTTCTTG	GCGAATACGT	
251	AATCCCATTT	TTTGTTTAGC	AAGAAAATGA	TCGGGATAAT	CATAATAGGT	
301	GTTGCCCAAA	AATAAATTTT	GATGTTCTAA	AATCATAAAT	GATGTTCTAA AATCATAAAT TTTGCAAGAT	33 /
351	ATTGTGGCAA	TTCAATACCT	ATTTGTGGCG	AAATCGCCAA	TTTTAATTCA ®	68
401	ATTTCTTGTA	GCATAATATT	TCCCACTCAA	ATCAACTGGT	TAAATATACA	
451	AGATAATAAA	AGATAATAAA AATAAATCAA	GATTTTTGTG	ATGACAAACA	ACAATTACAA	
501	CACCTTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAAAT	AGTATAAATC	
551	CGCCATATAA	AATGGTATAA	TCTTTCATCT	TTCATCTTTC	ATCTTTCATC	
601	TTTCATCTTT	CATCTTTCAT	CTTTCATCTT	TCATCTTTCA	TCTTTCATCT	
651	TTCATCTTTC	ATCTTTCATC	TTTCATCTTT	CACATGAAAT	GATGAACCGA	
701	GGGAAGGGAG	GGAGGGGCAA	GAATGAAGAG GGAGCTGAAC GAACGCAAAT	GGAGCTGAAC	GAACGCAAAT	
751	GATAAAGTAA	TTTAATTGTT	TTTAATTGTT CAACTAACCT	TAGGAGAAAA	TATGAACAAG	

#### FIG. 7B

TTGCCAAAGG	GGCGATATTT	GGTCAATCTG	AAAATGAAGC	GCCGCGCCTG	1551
TTACAGCATT	CAACCATTAC	ATAATAAACC	CATCAGCGAT	AAAAAATCAC	1501
CTCGCAGGGC	CATTTCTTTA	ATGGTGGCAG	ATTAGCGTAA	CGAGGGTGTG	1451
AAGTGAAAAA	ATTGGTGGCA	TGTAAATCTT	AAGACGGCAG	ACTGTCGGTA	1401
CGGTTTAATT	TTGTGAATCA	CTCGCTGAAA	AGATAAAGCG	AGCAAACCAA	1351
TTCACCTTCG	GGCGCGTAAT	AAAACATCAA	ATTTCTAACG	TACGCTAGAC	1301
TTACGGCTTC	ACTAATGGCT	AATTATTAAC	GTAAAGACGC	ATCACAATAG	1251
CCCAAATGGT	TTTTAATCAA	GGACAAGTCT	AGATTCTAAC	AAGGGATTTT	1201
TCCCAATTAA	TAACCAAATC	GTGTTACATC	GTATTCAACC	CAACTCCGCC	1151
AAGAAAACAA	CAGTTTTTAC	TGAAATGGTG	TCGACCAAAA	CAATTTAACA	1101
TAATTGGAAA	ACGCTATCAT	AACAGTGTTG	CATTATCCGC	GTAATAAAAC	1051
AAGAAAACAA	CAGTTTTTAC	TGAAATGGTG	TCGACCAAAA	CAATTTAACA	1001
TAGCAAGCGG	CAATCTGTTT	ATCTATTCCA	TAGGTGTAAC	TTACTATCTT	951
TTCCGCTATG	TAAAGCCACT	CACTTAGCGT	GAAAGTGCGT	CTGCTCGCAT	901
AGCGAAAAAC	AGAAAAAGGC	ACCATTCCAC	CGGGGTTGTG	TGAATTGGCA	851
TTGCTGTGTC	AATGCTTTGG	CAAACGCCTG	TCAAATTCAG	ATATATCGTC	801

#### FIG. 7C

CGGTAACA	TT	AATGTCCGTG	CGGTAACATT AATGTCCGTG CTGCCACTAT TCGAAACCAA GGTAAACTTT	TCGAAACCAA	GGTAAACTTT
CTGCTGATTC TGTAAC	TGTAAC	SCAAA	TGTAAGCAAA GATAAAAGCG GCAATATTGT	GCAATATTGT	TCTTTCCGCC
AAAGAGGGTG AAGCGGAAAT		AAAT	TGGCGGTGTA	ATTTCCGCTC	AAAATCAGCA
AGCTAAAGGC GGCAAGCTGA		TGA	TGATTACAGG	CGATAAAGTC ACATTAAAAA	ACATTAAAAA
CAGGTGCAGT TATCGACCTT	TATCGAC	CTT	TCAGGTAAAG	TCAGGTAAAG AAGGGGGAGA AACTTACCTT	AACTTACCTT
GGCGGTGACG AGCGCGG	AGCGCGG	CGA	AGCGCGCGA AGGTAAAAAC GGCATTCAAT	GGCATTCAAT	TAGCAAAGAA
AACCTCTTTA GAAAAAGGCT	GAAAAAG	3GCT	CAACCATCAA	TGTATCAGGC AAAGAAAAAG	AAAGAAAAG
GCGGACGCGC TATTGTGTGG	TATTGT	TGG	GGCGATATTG	CGTTAATTGA	CGGCAATATT
AACGCTCAAG GTAGTGGTGA	GTAGTGG	TGA	TATCGCTAAA	ACCGGTGGTT	TTGTGGAGAC ®
ATCGGGGCAT TATTTATCCA	TATTTAT(	CCA	TTGACAGCAA	TGCAATTGTT AAAACAAAAG	AAAACAAAAG
AGTGGTTGCT AGACCCT	AGACCCT	GAT	AGACCCTGAT GATGTAACAA	TTGAAGCCGA	AGACCCCCTT
CGCAATAATA CCGGTATAAA	CCGGTAT	AAA	TGATGAATTC	CCAACAGGCA	CCGGTGAAGC
AAGCGACCCT AAAAAAATA	AAAAAA	ATA		GCGAACTCAA AACAACGCTA ACCAATACAA	ACCAATACAA
CTATTTCAAA TTATCTG	TTATCTG	AAA	TTATCTGAAA AACGCCTGGA CAATGAATAT AACGGCATCA	CAATGAATAT	AACGGCATCA
AGAAAACTTA CCGTTAA	CCGTTAA	rag	CCGTTAATAG CTCAATCAAC	ATCGGAAGCA	ACTCCCACTT
AATTCTCCAT AGTAAAGG	AGTAAAG	FLC	AGTAAAGGTC AGCGTGGCGG AGGCGTTCAG	AGGCGTTCAG	ATTGATGGAG
ATATTACTTC TAAAGGC	TAAAGGC	GGA	TAAAGGCGGA AATTTAACCA TTTATTCTGG CGGATGGGTT	TTTATTCTGG	CGGATGGGTT

#### FIG. 7D

2451	GATGTTCATA	AAAATATTAC	GCTTGATCAG	GATGTTCATA AAAATATTAC GCTTGATCAG GGTTTTTTAA ATATTACCGC	ATATTACCGC
2501	CGCTTCCGTA	GCTTTTGAAG	GTGGAAATAA	GCTTTTGAAG GTGGAAATAA CAAAGCACGC GACGCGGCAA	GACGCGGCAA
2551	ATGCTAAAAT	TGTCGCCCAG	GGCACTGTAA	TGTCGCCCAG GGCACTGTAA CCATTACAGG AGAGGGAAAA	AGAGGGAAAA
2601	GATTTCAGGG	CTAACAACGT	ATCTTTAAAC	ATCTTTAAAC GGAACGGGTA AAGGTCTGAA	AAGGTCTGAA
2651	TATCATTTCA	TCAGTGAATA	ATTTAACCCA	ATTTAACCCA CAATCTTAGT	GGCACAATTA
2701	ACATATCTGG	GAATATAACA	ATTAACCAAA	CTACGAGAAA GAACACCTCG	GAACACCTCG
2751	TATTGGCAAA	CCAGCCATGA	TTCGCACTGG	TTCGCACTGG AACGTCAGTG	CTCTTAATCT
2801	AGAGACAGGC	GCAAATTTTA	CCTTTATTAA	ATACATTTCA	AGCAATAGCA W
2851	AAGGCTTAAC	AACACAGTAT	AGAAGCTCTG	CAGGGGTGAA	TTTTAACGGC
2901	GTAAATGGCA	ACATGTCATT	CAATCTCAAA	CAATCTCAAA GAAGGAGCGA	AAGTTAATTT
2951	CAAATTAAAA	CCAAACGAGA	ACATGAACAC AAGCAAACCT	AAGCAAACCT	TTACCAATTC
3001	GGTTTTTAGC	CAATATCACA	CAATATCACA GCCACTGGTG GGGGCTCTGT	GGGGCTCTGT	TTTTTTGAT
3051	ATATATGCCA	ACCATTCTGG	ACCATTCTGG CAGAGGGGCT	GAGTTAAAAA	TGAGTGAAAT
3101	TAATATCTCT	AACGGCGCTA	ATTTTACCTT	AAATTCCCAT	GTTCGCGGCG
3151	ATGACGCTTT	TAAAATCAAC	TAAAATCAAC AAAGACTTAA	CCATAAATGC AACCAATTCA	AACCAATTCA
3201	AATTTCAGCC	TCAGACAGAC	TCAGACAGAC GAAAGATGAT	TTTTATGACG GGTACGCACG	GGTACGCACG

#### FIG. 7E

-	<i>r</i>	<b>~</b>	<b>~</b>	_	rv		7/6		<b>~</b> 1	<u>د</u>	<b></b> 1	<i>.</i>	A	J.	Ph
GGTAA'I'G'I'C	TATTACTATC	CTAATCAGCA	GTTAATGGGA	TCTCACTATT	TAAATATCAC	ACACAAGGAG	CATTACCACT	TAATCAACAA	GAAATCCAAA	TTCTTCCGAT	TTGATGGAGA	ATTAAAACCA	CAATAAAGCA	ACAGTAATGA	AATGTTAAAG
CATTUTGGGC	TTACGGGGAA	AATAACGCCC	CAGCTTGCTC	TTAAAGGCAA	AGAGATACCC	AATTAATATA	GTGATTTAAA	GGCGGAGATA	TAATGATGCT	ACCTCACGAT	AAAAAGGGTA	CAACCTAACT	TTTCAGGTTT	ACTATTGGCA ACAGTAATGA	AACTTTTAAC AATGTTAAAG
AATTICAAUCI ACAACAIAIC CATTCTGGGC GGTAATGTCA	ACAAAACTCA AGCAGCAGCA	GCTAGAAGCC	TAAAACTTGG	AATGCAGATA	AGGAAAGACT	ACCAATAATG GCACTGCCGA AATTAATAA	ACCAATGATG	GCAACCAAAG AAGCATCATC	TTAAATATTA CAGACAGTAA	AAAGAAGGCA	TCACCAAACA GATAACAATC AAAAAGGGTA	TCAGATGCGA CAAGTAATGC	GACCTAAGTA	CCAAAGATGG TAGAGATTTA	GGTGCCGAAG CCAAAACAGT
AATICAACCI	ACAAAACTCA	CAAATGTTAC	GATAGAGTTA	AACTGGCGAA	CCACTTTTAA	ACCAATAATG	TGGCAATGTT	GCAACCAAAG	TTAAATATTA	TATCTCGCAA	TCACCAAACA	TCAGATGCGA	ATTGACAGAA GACCTAAGTA	CCAAAGATGG	GGTGCCGAAG
CAATGCCATC	CCCTTGGTGG	GAGAAAGCAG	AAACATAAGG	GTTTAAGTTT	TCAGAAAGCG	CGGCAATTTT	TGGTAAAACT	CACGCTAAAC	AAAAGGAAGC	TTGGCGGCAA	AAAATTAATA	GGACTCTAGT	AAGAATTGAA	GAGATTACAG	CGGTAACAGC
325I	3301	3351	3401	3451	3501	3551	3601	3651	3701	3751	3801	3851	3901	3951	4001

#### FIG. 7F

						3	8/8	86								
TAGCAAAGTG	ACAACGATAC	GATATTACTT	CACCACAGCA	CAACCAAAAC	GTTAGCGCGA	GAAATCGGGT	CAATTTCCGG	GTTGGGAATG	CGCAACAGGG	CTAAGGGTCA	ATTAATGCTG	GGCAGGCTCG	AAGATGCTAA	GCAGTCAACG	TGTGAATATC	CGAAAGATGG
ATTCAAAAAT CTCTGCTGAC GGTCACAATG TGACACTAAA TAGCAAAGTG	CGGACGTGAA AGCAATAGCG	ATTACTGCAA AAAATGTAGA AGTAAACAAA GATATTACTT	CTCTCAAAAC AGTAAATATC ACCGCGTCGG AAAAGGTTAC CACCACAGCA	TTAACGCAAC AAATGGCAAA GCAAGTATTA	CACGGTAAGT	AAATTGAAGC	AACAGGTACA ATTGGCGGTA	CAAACGCTGG CGATTTAACA GTTGGGAATG	GCGCAGAAAT TAATGCGACA GAAGGAGCTG CAACCTTAAC	ATCACTTCAA	CGCAGGAAGC	TAACCACCGT GGCAGGCTCG	ATTAACGCAA AAGATGCTAA	GTGATAGTAC AGAAGTGAAT	CCTCAAGCAG	TAAACACAGT AAATGGGTTA AATATCATTT
GGTCACAATG	CGGACGTGAA	AAAATGTAGA	ACCGCGTCGG	AAATGGCAAA	TTTCCGGTAA	TCCGGCTCAA	AACAGGTACA	CAAACGCTGG	GAAGGAGCTG	CGGTTCTAGC	ATGGTAGCAT	ACAGGCACCT	CACCTTGGTT	GTGATAGTAC	ACTGCGGCAA	AAATGGGTTA
CTCTGCTGAC	AAAACATCTA GCAGCAATGG		AGTAAATATC	TTAACGCAAC	AGGTGATATC AGCGGTACGA	AACCACTAAA	TAACAAGTGC	TAATACGGTA AATGTTACGG	TAATGCGACA	AATACCTTGA CTACTGAAGC CGGTTCTAGC	GGTAGACCTC TTGGCTCAGA	CTAATGTGAC ATTAAATACT	GATATTAAAG CAACCAGCGG	GATGCATCAG	TGGTAGTGTG	TAAACACAGT
ATTCAAAAAT	AAAACATCTA	CGGCTTAACT	CTCTCAAAAC	GGCTCGACCA	AGGTGATATC	CTGGTGATTT	GAGGCTAATG	TAATACGGTA	GCGCAGAAAT	AATACCTTGA	GGTAGACCTC	CTAATGTGAC	GATATTAAAG	GCTAAATGGT GATGCATCAG	ACTGGGGATT	ACTGGGGATT
4051	4101	4151	4201	4251	4301	4351	4401	4451	4501	4551	4601	4651	4701	4751	4801	4851

#### FIG. 7G.

4901	TAGAAACACT	GTGCGCTTAA	GAGGCAAGGA	TAGAAACACT GTGCGCTTAA GAGGCAAGGA AATTGAGGTG AAATATATCC	AAATATATCC
4951	AGCCAGGTGT	AGCAAGTGTA	GAAGAAGTAA	TTGAAGCGAA ACGCGTCCTT	ACGCGTCCTT
5001	GAAAAAGTAA	GAAAAAGTAA AAGATTTATC	TGATGAAGAA	TGATGAAGAA AGAGAAACAT	TAGCTAAACT
5051	TGGTGTAAGT	GCTGTACGTT	TTGTTGAGCC	TTGTTGAGCC AAATAATACA	ATTACAGTCA
5101	ATACACAAAA	TGAATTTACA	ACCAGACCGT	CAAGTCAAGT	GATAATTTCT
5151	GAAGGTAAGG	CGTGTTTCTC	GAAGGTAAGG CGTGTTTCTC AAGTGGTAAT GGCGCACGAG	GGCGCACGAG	TATGTACCAA
5201	TGTTGCTGAC	TGTTGCTGAC GATGGACAGC	CGTAGTCAGT	AATTGACAAG	GTAGATTTCA W
5251	TCCTGCAATG	AAGTCATTTT	ATTTTCGTAT	TATTTACTGT	GTGGGTTAAA 0
5301	GTTCAGTACG	GTTCAGTACG GGCTTTACCC		ATCTTGTAAA AAATTACGGA GAATACAATA	GAATACAATA
5351	AAGTATTTT	AACAGGTTAT	TATTATGAAA	AATATAAAA	GCAGATTAAA
5401	ACTCAGTGCA	ATATCAGTAT	TGCTTGGCCT	TGCTTGGCCT GGCTTCTTCA	TCATTGTATG
5451	CAGAAGAAGC	CAGAAGAAGC GTTTTTAGTA AAAGGCTTTTC	AAAGGCTTTC	AGTTATCTGG	TGCACTTGAA
5501	ACTTTAAGTG	AAGACGCCCA	ACTGTCTGTA	GCAAAATCTT	TATCTAAATA
5551	CCAAGGCTCG	CAAACTTTAA	CCAAGGCTCG CAAACTTTAA CAAACCTAAA AACAGCACAG	AACAGCACAG	CTTGAATTAC
5601	AGGCTGTGCT	AGATAAGATT	GAGCCAAATA	AATTTGATGT	GATATTGCCG
5651	CAACAAACCA	TTACGGATGG	CAATATCATG	TTACGGATGG CAATATCATG TTTGAGCTAG TCTCGAAATC	TCTCGAAATC

#### FIG. 7H

GATTTATTCT	TAGCAGTATA	TACAAGATAT	CAATTTACTC	ATTATCAGGT	6601
TTAGCAGTCA	GGTTGGCATT	GTTTGCTCAA	TGAGTCAAGA	AGTTTAGGGT	6551
TAGCACAGCC	GCTATCACAT	TTTAATCGCA	TGGCGAAACA	TGGAGCGCAT	6501
TCTTTTGGAA	ATTACCAGGC	ACGCGAGTAA	CATCATTATT	TGATTTAACT	6451
TCTTTAATAT	CCTAAAACAA	CCAATTTACC	ATGGACATAT	GCAGGCATTG	6401
AGGCGTAAGT	TTGCAGTATC	AAGAAAAAT	GGGTGAAACG	TAAATCGCTT	6351
ACCTCCGCGT	TATTAATCAA	ACTACCGCCA	TTAGGCTACA	TAAAATTAAT	6301
AAGACCAATT	CTTGGCATGG	AACATTTAAC	ATTATCTCCC	AAATGGAGTT	6251
TGCGAATCTG	AATCTATCTC	TCAAAAGGTC	TCGTAAATTA	GTGCGATTAA	6201
GGCTTACCAA	TGATATCGAC	CTGATTCTAA	ATGAGTTATG	TTATACCAGT	6151
GATGTGTTAA	TGGTCATGAT	CCAATTTAAC	TTTGTTAATG	AAGCTTGGGT	6001
ACCAACGTGT	GAGTTTAACT	CGGCGCGAGA	ATGATAATTT	TTTATTTCTT	5951
AACGCGTAGC	CTTTTGGTAA	GGCTTCTCGC	GATAATTGCG	CCTCTAATTT	5901
AAAAACAAAA	ACTAAACCCT	TACATTACGA	GTTACCCGTG	CCCGCTTAAG	5851
CAAAAGAAAA	TTTAATATGG	TTTGCGTGAA	AGTGGTTCGA	GATGGTCGTC	5801
AGTGTATGAA	AACAAGGAAA	CCATCTTTGA	TCGTAGCCTG	AAAATATCGC	5751
TATAGTGAAG	GAGCCAGGGT	TTTATAAGGC	AGCCAAGTTT	AGCCGCAGAA	5701

#### FIG.7I

6651	CTGTAACAGG	TACTTATGGC	TACTTATGGC GTCAGAGGCT	TTAAATACGG CGGTGCAAGT	CGGTGCAAGT
6701	GGTGAGCGCG	GTCTTGTATG	GTCTTGTATG GCGTAATGAA	TTAAGTATGC	CAAAATACAC
6751	CCGCTTCCAA	ATCAGCCCTT	ATGCGTTTTA	TGATGCAGGT	CAGTTCCGTT
6801	ATAATAGCGA	AAATGCTAAA	AAATGCTAAA ACTTACGGCG AAGATATGCA	AAGATATGCA	CACGGTATCC
6851	TCTGCGGGTT	TAGGCATTAA	TAGGCATTAA AACCTCTCCT	ACACAAAACT	TAAGCCTAGA
6901	TGCTTTTGTT	GCTCGTCGCT	TTGCAAATGC	CAATAGTGAC	AATTTGAATG
6951	GCAACAAAAA	ACGCACAAGC		TCACCTACAA CCTTCTGGGG GAGATTAACA	GAGATTAACA &
7001	TTCAGTTTCT	AACCCTGAAA	TTTAATCAAC	TGGTAAGCGT	TCCGCCTACC @
7051	AGTTTATAAC	TATATGCTTT	ACCCGCCAAT	TTACAGTCTA	TAGGCAACCC
7101	TGTTTTTACC	CTTATATATC	CTTATATATC AAATAAACAA GCTAAGCTGA		GCTAAGCAAA
7151	CCAAGCAAAC	TCAAGCAAGC	CAAGTAATAC	TAAAAAACA	ATTTATATGA
7201	TAAACTAAAG	TATACTCCAT	TATACTCCAT GCCATGGCGA	TACAAGGGAT	TTAATAATAT
7251	GACAAAAGAA	AATTTGCAAA	AATTTGCAAA ACGCTCCTCA	AGATGCGACC	GCTTTACTTG
7301	CGGAATTAAG	CAACAATCAA	CAACAATCAA ACTCCCCTGC	GAATATTTAA ACAACCACGC	ACAACCACGC
7351	AAGCCCAGCC	TATTACGCTT	TATTACGCTT GGAACAACAT	ATCGCAAAAA AAGATTATGA	AAGATTATGA
7401	GTTTGCTTGT	CGTGAATTAA	TGGTGATTCT	CGTGAATTAA TGGTGATTCT GGAAAAAATG GACGCTAATT	GACGCTAATT

#### FIG. 7J.

드	<u>-</u>	<u></u>	r s	<b>ب</b> س	ح.	42	2/68	3		<b>F</b> .		_	<b>.</b> .		_
CTGGCATA1	CTAATGCAAT	GAAGGGGCGT	TGCCTCTTCC	TCAACCCAGA	ATTGCTAAAT	TTTAGATGCG	TTGCGTTGCA Ø	AGAGCGGTGG	TTTAGATGAA	GTTATGATTT	CTTGTCCGCA	CACCTTAGGT	ATTTAATTC	GCTGCTCGAG	TAAAATAGGT
CACCCGCTCA GCTGGCATAT	ACTCGTCTCG C	AATTTCTGAA G	CGCTGATTTT T	AAATATATA T	CAACTCTTCT A	TCAATATGAG T	TCATTGTGTT T	GTTTCATAAA A	AAATTGCTAA T	ATGCACTGCA G	TTAAGCGTCC ATTAAACGAA C	GCTACCTTTA CA	CTGCTTGAAC A	TTCAATGATT GO	AGGGCGTTGA TA
GAATTTGACG	TTATTTGCC	CCGAATTGGC	CGCTGGTTGA	TATTCTCAAT	TAGCAACAGA	GAATCCAATG	ACTTTGTGCT	CCGCATCTGC	AAACTCGCCG	TGATGTATAT	TTAAGCGTCC	TGGCAAGACC	GATGATGGTA	CACATTCAAC	TTAGGCCATG
TCACGATATT	AATTACTAAT	TTTTCCGACC	TAGCCTGCAA	ACGCAGACCA	GGCTTTCATT	TTACTTACCC	GGAATCAACA	TTTATTGGTA	GTTTCCTAAA AAACTCGCCG	ATATCCTTCA	AAGCACGATG	CACGCAAGGA	GCAAACCTGT	ATTTATCGTA	AAAAATTCTA TTTAGTCGGC TTAGGCCATG
TTGGAGGCGT	CTACCCGAAA	TACAACACTC	TAAAGATGAT	CCCTACGTTA	TTCCGAAGGT	TCTGTATTTT	TTATGGGCAG	GTCTTCACGT	TTTTACAGTG	TTGCCTGCAA	AGCAAAAAC AAGCACGATG	AGCATATCCT CACGCAAGGA	AAAAAGGACG	GGGACATTCG	AAAAATTCTA
7451	7501	7551	7601	7651	7701	7751	7801	7851	7901	7951	8001	8051	8101	8151	8201

#### FIG. 7K

GTGT CAAC	CAAC		CTAC	GGCA	CCTA			TAAA 00	CACA	ACTG	TTGC	TTGA	GTAC	ATGG	TAGC	AACĄ
	CCCGCA	TGTGAGCAAC	CTGCCACTAC	TATGTGGGCA	AGATGCCCTA	ATGTACTCAG	ACAATGAAAT	TAAAGCTAAA	GCTTGACACA	GATGCC	GCGTGATTGC	GCATAATTGA	GATGAAGTAC	ACCAGAATGG	TGCGTCTAGC	ATAGAAAACA
	AACAGTGCGA AACTTTCCAA CCCGCAGTGT	CCACGATTTT	CTGGGTCATC	AGAAGATGAT	GCTTACCCAA	CGCCCCACAA AAAGTGGATT	TGCCGCTACC	AAATCAGAGA	CAATCAACAG	TTTAGGTGAC GATGCCACTG	TGGCAATATT	AATACTAACG	CAAAACGGGG	GCTTAGGACT	GAATGTGCTT	TCGTTACATC
	AACAGTGCGA	AAGCATTGGC ATGGATATTA	AGCTGTAGCC	ATGTCATCGT	TITCAGCGAA ACCCTITITAC	CGCCCCACAA	ATATCGGTAT	ACATTGCAAG	CGCACTTGGA	TCGAAAGCTA	CACGATTATC	TCCTTTCGGT	TTGGTGTATG	CTGTTTAAAC		TTGAACTCCG
	TTTATCCGTA	AAGCATTGGC	CCCCTATTCA	TTTATTGATT	TTTCAGCGAA	CTTCTGCACT	GAAGTAGTCA	ATTTTGCTA	ATTTTCATTT	AAATGGTTTA	CACATCCCCA CGCACCTTAT	TAAATCCGTT	TTAGGTTTAG	TGATGAAGGT	CTGATAGCCG ACACACGAGA AACATATTT	CAAGAACGCC
	GAGACTGTTT	TCTATATGCC	ACTCGGCTTG	GCATTCTGAA	GTGAAGATTG	CCTTATGTAC	GGAAAACCCT	TAAACCCTGA	GTCAAAATAC	CCCTTATGTC	CACATCCCCA	GATATGCTAC	TATGGTTACA	ATGAACATAT	CTGATAGCCG	AGAAAACCAT
	8301	8351	8401	8451	8501	8551	8601	8651	8701	8751	8801	8851	8901	8951	9001	9051

44/68

# FIG.7L

		AAA	AAATCACCAA TACCCACAAA AAA	AAATCACCAA	9401
TTGCACCACA	TAGCCAAAAC TGGCAGAAAT TAAAGGCTAA AATCACCAAA TTGCACCACA	TAAAGGCTAA	TGGCAGAAAT	TAGCCAAAAC	9351
GCGGAGATTT	CAGTTTATCA GCCTCCCGCC ATAAAACTCC GCCTTTCATG GCGGAGATTT	ATAAAACTCC	GCCTCCCGCC	CAGTTTATCA	9301
CGCACGCTGA	CTCTCAAAAA TCAACCGCAC TTTTATCTTT ATAACGATCC CGCACGCTGA	TTTTATCTTT	TCAACCGCAC	CTCTCAAAAA	9251
TTTTAAAAAC	ACGGTTTTTT AAAGTAAAAG TGCGGTTAAT TTTCAAAGCG TTTTAAAAAC	TGCGGTTAAT	AAAGTAAAAG	ACGGTTTTTT	9201
GTAAAAATA	CTGCTTAAGA AAACAAATGA ATGGAAGCGG AAGCACTTGA GTAAAAATA	ATGGAAGCGG	AAACAAATGA	CTGCTTAAGA	9151
GGGCAAAATA	CA AAAGCTTTTT ACAGGCGACC CTCGTCCATT GGGCAAATA	ACAGGCGACC	AAAGCTTTTT	ACGGCTTACA	9101

#### FIG. 8A

-	GATCAATCTG	GGCGATATTT	TTGCCAAAGG	GATCAATCTG GGCGATATTT TTGCCAAAGG TGGTAACATT AATGTCCGCG	AATGTCCGCG
51	CTGCCACTAT	TCGCAATAAA	GGTAAACTTT	CTGCCACTAT TCGCAATAAA GGTAAACTTT CTGCCGACTC TGTAAGCAAA	TGTAAGCAAA
101	GATAAAAGTG	GTAACATTGT	TCTCTCTGCC	TCTCTCTGCC AAAGAAGGTG AAGCGGAAAT	AAGCGGAAAT
151	TGGCGGTGTA		AAAATCAGCA	ATTTCCGCTC AAAATCAGCA AGCCAAAGGT GGTAAGTTGA	GGTAAGTTGA
201	TGATTACAGG	CGATAAAGTT	TGATTACAGG CGATAAAGTT ACATTGAAAA CGGGTGCAGT	CGGGTGCAGT	TATCGACCTT
251	TCGGGTAAAG	AAGGGGGAGA	AACTTATCTT	TCGGGTAAAG AAGGGGGAGA AACTTATCTT GGCGGTGACG	AGCGTGGCGA
301	AGGTAAAAAC	AGGTAAAAC GGCATTCAAT	TAGCAAAGAA	AACCACTTTA	GAAAAAGGCT A
351	CAACAATTAA	TGTGTCAGGT	AAAGAAAAAG	CAACAATTAA TGTGTCAGGT AAAGAAAAAG GTGGGCGCGC	AAAGAAAAA GTGGGCGCGC TATTGTATGG O
401	GGCGATATTG	CGTTAATTGA	CGGCAATATT	CGTTAATTGA CGGCAATATT AATGCCCAAG GTAAAGATAT	OTAAAGATAT
451	CGCTAAAACT	CGCTAAAACT GGTGGTTTTG	TGGAGACGTC	TGGAGACGTC GGGGCATTAC TTATCCATTG	TTATCCATTG
501	ATGATAACGC	AATTGTTAAA	ACAAAAGAAT	ATGATAACGC AATTGTTAAA ACAAAAGAAT GGCTACTAGA CCCAGAGAAT	CCCAGAGAAT
551	GTGACTATTG	GTGACTATTG AAGCTCCTTC	CGCTTCTCGC	CGCTTCTCGC GTCGAGCTGG GTGCCGATAG	GTGCCGATAG
601	GAATTCCCAC	TCGGCAGAGG	TGATAAAAGT	GAATTCCCAC TCGGCAGAGG TGATAAAAGT GACCCTAAAA AAAAATAACA	AAAAATAACA
651	CCTCCTTGAC	CCTCCTTGAC AACACTAACC AATACAACCA		TTTCAAATCT	TCTGAAAAGT
701	GCCCACGTGG	TGAACATAAC	GGCAAGGAGA	TGAACATAAC GGCAAGGAGA AAACTTACCG TTAATAGCTC	TTAATAGCTC
751	TATCAGTATA	GAAAGAGGCT	CCCACTTAAT	TATCAGTATA GAAAGAGGCT CCCACTTAAT TCTCCACAGT GAAGGTCAGG	GAAGGTCAGG

WO 94/21290 PCT/US94/02550

46/68

#### FIG.8B

GCAATATAAC	GCCACCGGTG	AAACTTAAAC	AAAATGATTT	TTTTCAATAG	1601
AGAAACCGCT	AAGGTTCAAC	CTCAAGGCTG	AACTTTAAAT	CAGGAGGGTC	1551
TTTAATGTCI	ATCTCAAAAC	TAATTATAAA	ACCCCTGGCG	CAACATACAA	1501
CCTCATCTAG	AAACTTAACG	CGTTAATTTC	GGGGGGGTAG	TCAGTCTCAG	1451
TGAAGATATT	CATTATTTAA	GCTAACTACG	TAAGAGTAAC	TAATGCCCTT	1401
AAGGCATCAA	CTTTAGCATC	CAACAGCTAA	GCACAAGGCT	TTTTAATATC	1351
ATAAAGCCAC	ATAACATTTA	ATTAAATGGC	GCAATGCAGA	CCAAGCATAC	1301
CTCAACAGGT	CAGGAAGTGG	ATTGACAGCA	TAACCTCTCC	GTAGTAAATT	1251
GTTACCTCGG	CACTTTAAAT	GGAACGTAAC	CGCACCTACT	AGACAAAGGA	1201
GGTTTTACAG	AAAGTCAGCT	GAAAGCACCC	ATATCTCAAT	GGAACTGTAG	1151
AAACATTTCC	ACGGAACGTT	AACAAATTTG	TAATATCTCA	GAACTAAGGG	1101
AGAGGTAGAA	CAGAGAGGAC	TTACTGACAG	AAGCTGAGCT	CCTTGGCGGA	1051
CTCTAAACAG	AACAACGTCT	CTTTAGATTT	ATAGTAACGG	ACCTCAGGTA	1001
AGGGACCATC	TTACAGCCCA	AACCTAACCA	TGGACGGAAC	AAGACAAGTC	951
ATCGCCTTCG	AGAAGGAGAT	TCACAACTAA	TTTTAAACA	TGGTAGCGGC	901
ATATTACGCT	GTTCATAAAA	ATGGGTTGAT	ATTCTGGCGG	TTAACCATTT	851
AGGCGGAAAT	TTACTTCTGA	GATAAAGATA	TGTTCAGATT	GCGGTCAAGG	801

### .IG.8C.

	本田ででであるが出て、本本で本でもで田本本		i i i i i i i i i i i i i i i i i i i	
) (1.5	<i>s</i> AGGGT'A	CCGATTCACG	CCGATTCACG CGTCAACAAA GGTGTCGCAG	GGTGTCGCAG
CAT	CCAAAAAAA CATAACTTTT	AAAGGGGGTA	ATATCACCTT	CGGCTCTCAA
CAG	AAATCAA	AAAGCCACAA CAGAAATCAA AGGCAATGTT	ACCATCAATA	AAAACACTAA
CGT(	CGTGGTGCGA	ATTTTGCCGA	AAACAAATCG	CCTTTAAATA
TGT	TGTTATTAAT	AATGGCAACC	TTACCACTGC	CGGCTCCA'I'T
ງຄວວ	ATCAATATAG CCGGAAATCT	TACTGTTTCA	AAAGGCGCTA	ACCTTCAAGC
TACA	TACACTTTTA	ATGTAGCCGG	CTCATTTGAC	AACAATGGCG
TTCC	TTCCATTGCC	AGAGGAGGGG	CTAAATTTAA AGATATCAAT	AGATATCAAT
GCTT	AACACCAGTA GCTTAAATAT	TACCACCAAC	TCTGATACCA	CTTACCGCAC
GGCA	CATTATAAAA GGCAATATAT	CCAACAAATC	AGGTGATTTG	AATATTATTG
CGAC	CGACGCTGAA	ATCCAAATTG	GCGGCAATAT	CTCACAAAAA
TCAC	TCACAATTTC	TTCTGATAAA	GTAAATATTA	CCAATCAGAT
GCAG	AACAATCAAA GCAGGCGTTG	AAGGGGGGCG	TTCTGATTCA	AGTGAGGCAG
CCTA	CCTAACTATT	CAAACCAAAG	AGTTAAAATT	GGCAGGAGAC
CAGG	CAGGCTTTAA	TAAAGCAGAA	ATTACAGCTA AAATGGCAG	AAATGGCAG
ATTG	ATTGGCAATG	CTAGCGGTGG	TAATGCTGAT GCTAAAAAG	GCTAAAAAAG

#### FIG. 81

ATAGTGCAAA	ACTATTGGAA	TGGTGATTTA	CAGCAAGCAC	GTAAATGTTA	3251
TGGTAATACA	GTACAATTTC	GATATTGAAG	CCAATCAGGC	CCACCTCAAG	3201
AATAGTGTAA	TAATGGGACT	GTTCTACAAT	TCCACAGTAG	TAAATTAACC	3151
CGGATAGCGG	ACTATTACTG	TAACACTGTT	ATATTTCAGG	GCTGTAGGTA	3101
AGCAACTCTT	TTGCAACTGG	GTAACACTTG	CTCCGGCTCT	TTGAATCCAG	3051
AACGGTAAAG	AGGTGATATC	CAACCAAAAC	GCAAATATTA	AACAGGCAAT	3001
TTAGTGCGAC	GGCTCAACCA	AACTACAGCA	GAGCCTTGAC	GCGGATGCAG	2951
AACAGTAACA	GTCAAGATGT	AATATCACTG	TAAGGTAAGT	GCAATACACT	2901
ACAGCGAGCG	TGTAAATATT	CTTCCGGTAA	ATTGAATCAA	TAAAGGTGGA	2851
CAGGGGATAT	AGTACAAAAA	AGTAAACATT	CCAGCGGCAC	ATTAATGCAA	2801
GAATGCTGTC	TTACCACAGA	GAAAATCTTG	GACAGCAACA	ATGTAACAGT	2751
ACCTCGCAAA	AGGCAACATT	GTACAATTAA	GCTCAAAATG	GGAAGTAACT	2701
CAGGCAGCGT	AATGCAACCA	CACAACTATC	CCAAAGAAGG	AATGTAACAA	2651
CGCAGCAGGA	ATATCTCTGC	AAGACAATAA	TACCTCCCAC	ACAATAACGT	2601
GTAACGGTAA	CGCAAAAGAT	TAACCATTTC	AGCACCGGTT	TAATGATAAC	2551
GCAATGCTGG	AATGGTAGTA	GAAAACGTCT	ATAGCGAAGT	GTAACACTAA	2501
CGGTCACAAT	TCTCGACTGA	GATTCAAAAA	CAAGGTTAAA	TGACTTTTGA	2451

#### FIG. 8E

Ŧ	E	Ą	H	r D		ر)68 ن		r	A	A	A	H	Æ	E
TCAGACAACT	CTGCTAATGT	TCAAAGATTA	CAAATTAGAT	ACGCAAGTG	ATCACCGGGG	TGGTAGAAAC	TCCAACCAGG	CTTGAGAAGG	ACTTGGTGTA	TTAATACACA	TCTGAAGGTA	CAATGTTGCT	TCATCCTGCA	AAAGTTCAG
CAAGCAATGG	TATCGCAGGA AACATTAATG	TACAGGGGAT	CAAAAGATGC	AATGCAACTA	CAGCGTGAAT	TTTCGGAAAA	GTGAAATATA	GAAACGCGTC	CACTAGCCAA	GCCATTACGG	AGTGACAATT	GAGTATGTAC	AAGGTAGATT	TATTATTTAC TGTGTGGGTT AAAGTTCAGT
AACAGGCTCT AGCATTACCT CAAGCAATGG	TATCGCAGGA	ACCACAGGCA CTTTAACTAC TACAGGGGAT	TGGTACCTTA ACAATCAATG	CAGGTGACCG CACAGTAGTA AATGCAACTA ACGCAAGTGG	GTGACTGCGA AAACCTCAAG CAGCGTGAAT	TTAAATATCA	TAAGAGGCAA GGAAATTGAT	GTAGAAGAGG TAATTGAAGC GAAACGCGTC	ATCTGATGAA GAAAGAGAAA CACTAGCCAA	GTTTCGTTGA GCCAAATAAT	ACAACCAAAC CATCAAGTCA AGTGACAATT	AATGGCGCAC GAGTATGTAC	AGCAGTAGTC AGTAATTGAC AAGGTAGATT	TATTATTTAC
AACAGGCTCT	AGGATAGCAG	ACCACAGGCA	TGGTACCTTA	CAGGTGACCG	GTGACTGCGA	AATAAATGGG	TAAGAGGCAA	GTAGAAGAGG	ATCTGATGAA	GTTTCGTTGA	ACAACCAAAC	CTCAAGTGGT	AGCAGTAGTC	TTTATTTTCG
TAACCACCCA	CTTACAGCCA	GACGTTAAAT	ACGCAACCAG	GGTGCTGCAT	CTCTGGTAAC	ATTTAAACAC	ACTGTGCGCT	TGTAGCAAGC	TAAAAGATTT	AGTGCTGTAC	AAACGAGTTT	AGGCGTGTTT	GACGATGGAC	ATGAAGTCAT
3351	3401	3451	3501	3551	3601	3651	3701	3751	3801	3851	3901	3951	4001	4051

50/68

## FIG. 8F

CAATATCAA TATTGCTTGG CTTGGCTTCT TCATCGACGT ATGCAGAAGA	TCATCGACGT	CTTGGCTTCT	A TATTGCTTGG CTTGGCTTCT TCATCGA	GCAATATCAA	4201
i i i i i i i i i i i i i i i i i i i				スペン 日々日々 なりび	4201
AAAACTCAGT	TTTAACAGGT TATTATTG AAAAACATAA AAAGCAGATT AAAACTCAGT	AAAAACATAA	TATTATTATG	TTTAACAGGT	4151
ATAAAGTATT	A CCCACCTTGT AAAAATTAC GAAAAATACA ATAAAGTATT	AAAAAATTAC	CCCACCTTGT	ACGGGCTTTA	4101

### FIG.9A

					5	51/6	88								
GACGGCAATA	TCATCAATTG GAAACAATTT	TTACAAGAAA GCAGCAACTC	TTAAAAGGGA	TGGTATCACA	CTTCTACGCT	CTTGAGCAAA	AATTACCGTT	AAAACGAGGG	GGGCAAAAAA	CATTGCTGCA	AAGGTGGTAA	CTTTCTGCCG	TGCCAAAGAA	AGCAAGCCAA	AAAACAGGTG
GTCGTACACG GTACAGCAAC CATGCAAGTA GACGGCAATA	TCATCAATTG	TTACAAGAAA	CATCTGACCA AATCTCCCAA	TCAACCCAAA	TAACACTAAT GGCTTTACTG	TAATTTCACC	ATCACGGTTT	GGCAAAGTGA AAAACGAGGG	TTTACTTGCA	TCACTTACAG	ATTTTTGCCA AAGGTGGTAA	TAAAGGTAAA	TTGTTCTCTC	AAATTGGCGG TGTAATTTCC GCTCAAAATC AGCAAGCCAA	CAGGTGATAA AGTCACATTA AAAACAGGTG
GTACAGCAAC	CCGTAATAGC GTCAATGCTA	GGAGCAGTTT	CATCTGACCA	GTCTTTTTAA	TAACACTAAT	TCAAGGCGCG	GAAATCGTGA	CCTTATTGGT	GTAGTATTTĆ	AATCCAACCA	AAGCGATCAA TCTGGGCGAT	CTATTCGCAA	AGTGGTAACA	TGTAATTTCC	CAGGTGATAA
GTCGTACACG	CCGTAATAGC	AAAATGAAAT	AACCGTGTTA	TAACGGACAA	ACGCAATTAT	AACGAAAACA	AGCACTCGCT	GTAGCGTAAA	GTAAATGGCG	CGATATAATA	AAGCGATCAA	CGCGCTGCCA	CAAAGATAAA	AAATTGGCGG	TTGATGATTA
GGGAATGAGC	AAACCACTAT	AACATTGACC	TGCCGTTTTC	TTTTAGATTC	ATAGGTAAAG	AGACATTTCT	CCAAGGATAA	GGTAAAGACG	CGTGATTAGC	TCACCATCAG	CCTGAAAACG	CATTAATGTC	ACTCTGTAAG	GGTGAAGCGG	AGGTGGTAAG
$\leftarrow$	51	101	151	201	251	301	351	401	451	501	551	601	651	701	751

#### FIG.9B

801	CAGTTATCGA	CCTTTCAGGT	AAAGAAGGGG	CAGTTATCGA CCTTTCAGGT AAAGAAGGGG GAGAGACTTA TCTTGGCGGT	TCTTGGCGGT
	GATGAGCGTG	GATGAGCGTG GCGAAGGTAA AAATGGTATT	AAATGGTATT	CAATTAGCGA	AGAAAACCTC
	TTTAGAAAAA	TTTAGAAAA GGCTCGACAA	TTAATGTATC	AGGCAAAGAA AAAGGCGGGC	AAAGGCGGGC
	GCGCTATTGT	ATGGGGCGAT	ATTGCATTAA	TTAATGGTAA	CATTAATGCT
	CAAGGTAGCG	ATATTGCTAA	AACTGGCGGC	TTTGTGGAAA CATCAGGACA	CATCAGGACA
	TGACTTATCC	ATTGGTGATG	ATGTGATTGT	TGACGCTAAA GAGTGGTTAT	GAGTGGTTAT
	TAGACCCAGA	TGATGTGTCC	ATTGAAACTC	TTACATCTGG	ACGCAATAAT
	ACCGGCGAAA	ACCGGCGAAA ACCAAGGATA	TACAACAGGA	GATGGGACTA	GATGGGACTA AAGAGTCACC
	TAAAGGTAAT	TAAAGGTAAT AGTATTTCTA	AACCTACATT	AACAAACTCA	ACTCTTGAGC ®
	AAATCCTAAG	AAATCCTAAG AAGAGGTTCT	TATGTTAATA	TCACTGCTAA	TAATAGAATT
	TATGTTAATA	TATGTTAATA GCTCCATCAA	CTTATCTAAT GGCAGTTTAA		CACTTCACAC
	TAAACGAGAT	GGAGTTAAAA	TTAACGGTGA	TATTACCTCA	AACGAAAATG
	GTAATTTAAC	CATTAAAGCA	GGCTCTTGGG	TTGATGTTCA	TAAAAACATC
	ACGCTTGGTA	CGGGTTTTTT	GAATATTGTC	GCTGGGGATT	CTGTAGCTTT
	TGAGAGAGAG	TGAGAGAGA GGCGATAAAG	CACGTAACGC AACAGATGCT	AACAGATGCT	CAAATTACCG
	CACAAGGGAC	CACAAGGGAC GATAACCGTC AATAAAGATG	AATAAAGATG	ATAAACAATT	TAGATTCAAT
	AATGTATCTA	TTAACGGGAC GGGCAAGGGT	GGGCAAGGGT	TTAAAGTTTA	TTGCAAATCA

### FIG.9C

GAGGGGAATT	TATATCTGTT	CTCTTGGCAA	AGAACTCTAA	CTTGAAGAAA	2401
GCAACACAGG	ACCAGCAACA	ACAAGCTGAC	ATGTTACATT	AATAAAGCAA	2351
CAATATCACC	CGGGCAATAT	AGTAGCATTA	AAATTCAAGC	TAGGTGGGGA	2301
AATGTCACTC	TCTTGGCGGC	ATCTAACCAT	TCAAGTCATA	CGCCATTAAC	2251
ACAGCAAACA	TATAATGAAT	AGATTCTTTT	AGCAAACGAA	TTTAGTCTTA	2201
TGGCTCGAAT	TAAATGCAAC	GACTTAACTA	AATCAAAAAA	ATGCTTTTGA	2151
CGCAATAGTA	ATCCCATAAT	TTTCCATAAC	GGGCTTGACT	CATTACCGGC	2101
ATTCAATTAA	ATAAACATGG	AGCTGCCGGC	TTACCTCTAG	CACGCCAATC	2051
GTTTGACATA	GCTCTGTGAT	AACAGTGATA	AGCTACCGGT	CCAACATTAC	2001
ACTTTTAACG	ATTACCTATT	CAAAAAAAGA	GCTACAGACC	ACCAAACGCC	1951
TTAAATTAAA	AAAGCCTTAT	AGCTAACGCA	TCAACATCGG	AAAACAAACT	1901
CATCGGAGGC	ATTTTAACGG	GCAGGCGTAC	TAGAAGTTTT	GGTCATCACG	1851
CAAGATTTGA	CTCAAATTCC	TTGATAGCGG	ATAAAATTCG	ATTTACCTTT	1801
CGGTGCAAAA	ACTTTGAATA	TTCTTCTTT	ACTGGAATGT	AAAGACTCTT	1751
GAATGCATCA	TTAAATACTG	AAAAAAGATG	CCAAACCACG	TAACAATTAA	1701
TCTGGAATAG	AATTAACATA	TTGATGGCGA	ACTCATAAAT	AAATAATTTC	1651

53/68

#### FIG.9D

ATGGTACAAT	ACTGCTCAAA	CGTGGAAGTA	CCACAGGCAG	ATCAATGCAA	3251
AGGCACAACT	CAACCAAAGA	GGAAATGTAA	TGCCGCAGCA	TAAATATCTC	3201
CACAAGACAA	CGTTACCTCC	TAAACAATAA	GATGTAACGG	TTCCGCAAAA	3151
GTTTAACCAT	AACAGCACCG	TGGTAATGAT	GTAGCAATGC	TCTAATGGTA	3101
AGTGAAAACG	TAAATAGCGA	AATGTAACAC	TGACGGTCAC	AAATCTCGAC	3051
AAAGATTCAA	TGACAAGGTT	AAGTGACTTT	GATGCTAAAA	TGGTAATGCT	3001
ATGCTAGCGG	ACTATTGGCA	CAGTGATTTA	CTAAAAATGG	GAAATTACAG	2951
TAATAAAGCA	TTTCAGGCTT	GACCTAAATA	ATTGGCAGGA	AAGAGTTAAA	2901
ATTCAAACCA	TAACCTAACT	CAGAAAATGC	TCAAGTGAGG	GCGTTCTGAT	2851
TTGAAGGGGG	AAAGCAGGCG	GATAACAATC	TTACCAATCA	AAAGTAAATA	2801
TTCTTCTGAT	ATCTCACAAT	AAAGAAGGCA	TATCTCACAA	TTGGCGGCAA	2751
GAAATCCAAA	AGCCGACGCC	AGAATATTAA	TTAAACATCA	AAAAGGCGAC	2701
TAACTAACGA	AACGGAAATA	AACCATTATT	GCACTCAAAA	AACGCCTCAG	2651
TATCACTACT	GTGGTTTAAA	ATCAATAAAG	AGGCGATATT	TAAAACTCCA	2601
CAAGGAGTGG	TAATATAAAA	CCGCCAACAT	AACAACGGTA	CACCTTTACC	2551
ACATCACCGG	GACAACCTAA	AGAAGCCAGT	CATTTAAAGG	GAAGATTCCA	2501
TTCTATTGCA	TCGGCAATCT	GCAAACATTG	TGGTGCAAAT	TAAGCCTAAC	2451

54/68

#### FIG. 9E

TTAACAATCA	TTAACGCAAC CAGTGGTACC TTAACAATCA	TTAACGCAAC	GATTCAAAGA	TACTACAGGG	4051
GCACTTTAAC	AATACCACAG	TGTGACGTTA	ATGCTGCTAA	GGAAACATTA	4001
CAGTATCGCA	CCAAGGATAG	ACTCTTACAG	TGGTCAGACA	CCTCAAGCAA	3951
TCTAGCATTA	CCAAACAGGC	AATTAACCAC	GAATCAGGCA	CTTAACTGCT	3901
GAGCTGCAAC	GCGAAAAATG	AAAAGTTGAA	GAAATAGTGC	TTAACTATTG	3851
CACTGGTGAT	TTACAGCAAG	ACAGTAAATG	TTCTGGTAAT	AAGGTACAAT	3801
GGCGATATTG	AAGCCAATCA	TAACCACCTC	ACTAATAGTG	AATTAATGGG	3751
TAGGTTCTAC	ACCTCCACAG	CGGTAAATTA	CTGCGGATAG	<b>GTTACTATTA</b>	3701
AGGTAACACT	GTAATATTTC	CTTGCTGTAG	TGGAGCAACT	TTGTTGCAAC	3651
TCTGTAACAC	CAGCTCCGGC	AAGTTGAATC	ATCAACGGTA	AACAGGTGAT	3601
TTACAACCAA	AATGCAAATA	GACAACAGGC	CCATTAGTGC	GCAGGCTCAA	3551
GACAACTACA	CAGGAGCCTT	ACAGCGGATG	TGTAACAGTA	CTGGTCAAGA	3501
AGTAATATCA	ACTTAAGGTA	GCGGCAATAC	ATTACAGCGA	TAATGTAAAT	3451
CAACTTCCGG	GGAATTGAAT	TATTAAAGGT	AAACAGGGGA	ATTAGTACAA	3401
CACAGTAAAC	CAACCAGCGG	GTCATTAATG	AGAGAATGCT	TTGTTACCAC	3351
ACAGAAAATC	AGTGACAGCA	AAAATGTAAC	ATTACCTCGC	TAAAGGCAAC	3301

55/68

### FIG. 9F

					TA	4701
	TGTAAAAAAT	TTACCCACCT	AGTACGGGCT	GTTAAAGTTC AGTACGGGCT	TACTGTGTGG	4651
	TCGTATTATT	CATTTTATTT	ATTTCATCCT GCAATGAAGT	ATTTCATCCT	GACAAGGTAG	4601
	GTCAGTAATT	GACAGCAGTA	GCTGACGATG	TACCAATGTT	CACGAGTATG	4551
	GGTAATGGCG	TTTCTCAAGT	GTAAGGCGTG	ATTTCTGAAG	TCAAGTGACA	4501
68	AACCATCAAG	TTTACAACCA AACCATCAAG	ACAAAACGAG	CGGTTAATAC	AATGCCATTA	4451
56/6	TGAGCCAAAT	TACGTTTCGT	GTAAGTGCTG	CAAACTTGGT	AAACACTAGC	4401
5	GAAGAAAGAG	TTTATCTGAT	GTCCTTGAGA AGGTAAAAGA	GTCCTTGAGA	AGCGAAACGC	4351
	AGGTAATTGA	ATATCCAACC AGGTGTAGCA AGCGTAGAAG AGGTAATTGA	AGGTGTAGCA	ATATCCAACC	GATGTGAAAT	4301
	CAAGGAAATT	AAATGGTAGA AACACTGTGC GCTTAAGAGG CAAGGAAATT	AACACTGTGC	AAATGGTAGA	TCATTTCGGA	4251
	GGGTTAAATA	AATATCACCG GGGATTTAAA CACAATAAAT GGGTTAAATA	GGGATTTAAA	AATATCACCG	AAGCAGCGTG	4201
	CGAAAACCTC	CTAACGCAAG TGGCTCTGGT AACGTGACTG CGAAAACCTC	TGGCTCTGGT	CTAACGCAAG	GTAAATGCAA	4151
	CCGCACAGTA	TGCCAAATTA GATGGTGCTG CATCAGGTGA CCGCACAGTA	GATGGTGCTG	TGCCAAATTA	ATGCAAAAGA	4101

# DERIVED AMINO ACID SEQUENCE FIG. 10A. COMPARISON OF

						57	/68							
SECOENCE	50	•	•	KVRHLALKPL	KVRHLALKPL		100	•	TIRNSVNAII	TIRNSVNAII		150	•	DSNGQVFLIN
DERIVED AMINO ACID SEQUENCE		•	•	EKGSEKPARM	EKGSEKPARM			•	ATMQVDGNKT	ATMQVDGNKT	ATMQVDGNKT		•	DQISQLKGIL
		•	•	ELARGCDHST	ELARGCDHST			•	GMSVVHGT	LQGMSVVHGT	LQGMSVVHGT		•	NSAVFNRVTS
		•	•	KRLNALVAVS	KRLNALVAVS			•	•	SIPQSVLASG	SIPQSVLASG		•	EMEQFLQESS
	₽	•	•	MNKIYRLKFS	MNKIYRLKFS		51	•	•	SAMLLSLGVT	SAMLLSLGVT	101		NWKQFNIDQN
, I i		Hmw3com	Hmw4com	Hmw1com	Hmw2com			Hmw3com	Hmw4com	Hmw1com	Hmw2com		Hmw3com	Hmw4com

INLGDIFAKG GNINVRAATI RNKGKLSADS VSKDKSGNIV

Hmw3com

### FIG. 10

Hmw1com Hmw2com	NWKQFNIDQN NWKQFNIDQN		EMVQFLQENN NSAVFNRVTS EMVQFLQENN NSAVFNRVTS	NQISQLKGIL	DSNGQVFLIN
	151				200
Hmw3com	•	•	•	•	
Hmw4com	PNGITIGKDA	IINTNGFTAS	TLDISNENIK ARNFTLEQTK	ARNFTLEQTK	DKALAEIVNH
Hmw1com	PNGITIGKDA	IINTNGFTAS	TLDISNENIK	ARNFTLEQTK	DKALAEIVNH G
Hmw2com	PNGITIGKDA	IINTNGFTAS	TLDISNENIK	ARNFTLEQTK	DKALAEIVNH 0
	201				250
Hmw3com		•	•	•	•
Hmw4com	GLITVGKDGS	GLITVGKDGS VNLIGGKVKN	EGVISVNGGS	ISLLAGORIT	ISDIINPTIT
Hmw1com	GLITVGKDGS	VNLIGGKVKN	EGVISVNGGS	ISLLAGQKIT	ISDIINPTIT
Hmw2com	GLITVGKDGS	VNLIGGKVKN	EGVISVNGGS	ISLLAGOKIT	ISDIINPTIT
	251				300

# FIG. 10C

RNKGKLSADS VSKDKSGNIV VSKDKSGNIV VSKDKSGNIV RNKGKLSADS RNKGKLSADS INLGDIFAKG GNINVRAATI VNLGDIFAKG GNINVRAATI GNINVRAATI VNLGDIFAKG YSIAAPENEA YSIAAPENEA YSIAAPENEA Hmw4com Hmw1com Hmw2com

59/68 350 IDLSGKEGGE IDLSGKEGGE IDLSGKEGGE IDLSGKEGGE GGVISAQNQQ AKGGKLMITG DKVTLKTGAV GGVISAQNQQ AKGGKLMITG DKVTLKTGAV GGVISAQNQQ AKGGKLMITG DKVTLKTGAV DKVTLKTGAV GGVISAQNQQ AKGGKLMITG LSAKEGEAEI LSAKEGEAEI LSAKEGEAEI LSAKEGEAEI 301 Hmw1com Hmw2com Hmw3com Hmw4com

400 TYLGGDERGE GKNGIQLAKK TTLEKGSTIN VSGKEKGGRA IVWGDIALID GKNGIQLAKK TTLEKGSTIN VSGKEKGGRA IVWGDIALID IVWGDIALID IVWGDIALID GKNGIQLAKK TTLEKGSTIN VSGKEKGGRA GKNGIQLAKK TTLEKGSTIN VSGKEKGGRA TYLGGDERGE TYLGGDERGE TYLGGDERGE 351 Hmw2com Hmw1com Hmw3com Hmw4com

# FIG. 10D.

450 SGHYLSIESN AIVKTKEWLL DPDDVTIEAE SGHYLSIDDN AIVKTKEWLL DPENVTIEAP SGHDLSIGDD VIVDAKEWLL DPDDVSIETL AIVDAKEWLL DPDNVTINAE SGHDLFIKDN GNINAQGK.D IAKTGGFVET GNINAQGS.D IAKTGGFVET IAKTGGFVET IAKTGGFVET GNINAQGSGD GNINAQGSGD 401 Hmw3com Hmw4com Hmw1com Hmw2com

/68 60 500 SASRVELGAD RNSHSAEVIK VTLKKNNTSL TTLTNTTISN LLKSAHVVNI ILRRGSYVNI ILKKGTFVNI SDPKKNSELK TTLTNTTISN YLKNAWTMNI PTLTNSTLEQ STPKRNKE.K TTLTNTTLES ESPKGNSISK TSGRNNTGEN QGYTTGDGTK DEYTGSGNSA DEFPTGTGEA TAGRSNTSED DPLRNNTGIN 451 Hmw3com Hmw4com Hmw1com Hmw2com

550 TLHTK...RD GVKINGDITS NE...NGNLT TLWSEGRSGG GVEINNDITT GDDTRGANLT .E. GGNLT · · · SKGGNLT SISIERGSHL ILHSEGQGGQ GVQIDKDITS ILHSKGQRGG GVQIDGDIT. SINLSNGS.L SINL. SNGSL SINGSNGSHL TARRKLTVNS TANORIYVNS TANNRIYVNS TASRKLTVNS 501 Hmw3com Hmw4com Hmw1com Hmw2com

# FIG. 10E.

	551				009	
Hmw3com	IYSGGWVDVH	KNITLGS.GF	IYSGGWVDVH KNITLGS.GF LNITTKEGDI AFEDKSGR	AFEDKSGR	NNLTITAO	
Hmw4com	IKAGSWVDVH	KNITLGT.GF	KNITLGT.GF LNIVAGDS.V AFEREGDKAR NATDAQITAQ	AFEREGDKAR	NATDAQITAQ	
Hmw1com	IYSGGWVDVH	KNISLGAQGN	INITAKQD.I	INITAKQD.I AFEKGSNQV.	ODLI	
Hmw2com	IYSGGWVDVH	KNITLD.QGF	KNITLD.QGF LNITA.AS.V AFEGGNNKAR DANNLTITAQ	AFEGGNNKAR	DANNLTITAQ	
	601				61/6 01/9	61/6
Hmw3com	GTITSG.NSN	GFRFNNVSLN	GTITSG.NSN GFRFNNVSLN SLGGKLSFTD SREDRGRRTK GNISNKFNGT	SREDRGRRTK	GNISNKFDGT	8

Hmw3com	GTITSG.NSN	GFRFNNVSLN	SLGGKLSFTD	SREDRGRRTK GNISNKFDGT	GNISNKFDGT	}
Hmw4com	GTITVNKDDK	QFRFNNVSIN	GTGKGLKFIA NON.	NON	.NFTHKFDGE	
Hmw1com	GTIT. SGNOK	GFRFNNVSLN	GTGSGLQFTT	GFRFNNVSLN GTGSGLQFTT KRTNK YAITNKFEGT	YAITNKFEGT	
Hmw2com	GTVTITGEGK	DFRANNVSLN	DFRANNVSLN GTGKGLNIIS	SVINI	LTHNLSGT	
	651				700	
Hmw3com	LNISGTVDIS	MKAPKVSWFY	RD.KGRTYWN	MKAPKVSWFY RD.KGRTYWN VTTLNVTSGS KFNLSIDSTG	KFNLSIDSTG	•
Hmw4com	INISGIVTIN	QTTKKDVKYW	NA.SKDSYWN	QTTKKDVKYW NA.SKDSYWN VSSLTLNTVQ KFTF.IKFVD	KFTF.IKFVD	
Hmw1com	LNISGKVNIS	MVLPKNESGY	DKFKGRTYWN	LTSLNVSESG EFNLTIDSRG	EFNLTIDSRG	

Hmw2com	INISGNITIN	TIN QTTRKNTSYW QTSHD.SHWN VSALNLETGA NFTF.IKYIS	QTSHD.SHWN	VSALNLETGA	NFTF. IKYIS	
	701				750	
Hmw3com	SGSTGPS	IRNAELNG	ITFNKA	TFNIAQGSTA NFSIKASIMP	NFSIKASIMP	
Hmw4com	SGSNSQD	LRSSRRSFAG	VHFNGIGGKT	NFNIGANAKA LFKLKPNAAT	LFKLKPNAAT	
Hmw1com	SDSAGTLTQ.	PYNLNG	ISFNKDT	TFNVERNARV NFDIKAPIGI	NFDIKAPIGI	
Hmw2com	SNSKGLTTQY	RSSAGVNFNG	RSSAGVNFNG VNGNM SFNLKEGAKV NFKLKPNENM	SFNLKEGAKV	NFKLKPNENM	62
						2/68
	751				800	
Hmw3com	FKSNANYAL.	FNEDISVSG.	.GGSVNFKLN	GGSVNFKLN ASSSNIQTPG VIIKSQNFNV	VIIKSQNFNV	
Hmw4com	DPKKELPIT.	FNANITATGN	SDSSVMFDIH	SDSSVMFDIH ANLTSRA AGINMDSINI	AGINMDSINI	
Hmw1com	NKYSSLNYAS E	FNGNISVSG. GGSVDFTLL ASSSNVQTPG VVINSKYFNV	GGSVDFTLL A	SSSNVQTPG V	VINSKYFNV	

850 SGGSTLNLKA EGSTETAFSI ENDLNLNATG GNITIRQVEG T. DSRVNKG SNFSLKQTKD SFYNEYSKHA TGGLDFSITS HNRNSNAFEI KKDLTINATG 801 Hmw3com Hmw4com

.GGSVFFDIY ANHS...GRG AELKMSEINI

NTSKPLPI.R FLANITATG.

Hmw2com

# FIG. 10G.

STGSSLRFKT SGSTKTGFSI EKDLTLNATG GNITLLQVEG T..DGMIGKG SNFSLRQTKD DFYDGYARNA HVRGDDAFKI NKDLTINATN SNGANFTLNS Hmw1com Hmw2com

900

851

63/68 ITNKANVTLQ ADTSNSNTGL VAAKKNITFK GGNITFGSQK ATTEIKGNVT INKNTNATLR GANFAEN.. SSSSITGNIN GGNVTLGGEN INSSHNLTIL Hmw3com Hmw4com

IEKAANVTLE ANNAPNQQNI GSDFDNHQ.. INNNANVTLI SSSSITGNIT GGNITFGSRK AVTEIEGNVT INSTYNISIL GGNVTLGGQN IVAKKNITFE Hmw2com Hmw1com

950 901

KSPLNIAGNV INNGNLTTAG SIINIAGNLT VSKGANLQAI TNYTFNVAGS SVEGNLSLTG ANANIVGNLS IAEDSTFKGE ASDNLNITGT TNFTFNVGGL NIVNIAGNLT VESNANFKAI INSGNLTAGG KKRTLTLGNI KPLTIKKDVI Hmw3com Hmw1com Hmw4com

TRDTLNITGN ISESATFKGK ENADIKGNLT RDRVIKLGSL LVNGSLSLTG Hmw2com

# FIG. 10H.

IIKGNISNKS IINGNITNEK IISGNITNKN IIGGDIINNK TTNSDTTYRT TTHAKRNQRS TTNSSSTYRT TTNASGTOKT IARGGAKFK. DINNTSSLNI ITQGVVKLG. NVTNDGDLNI DIDNSKNLSI DINNKGGLNI IAKGGARFK. IKQGVVKLQG FDNNGASNIS FTNNGTANIN FDNKGNSNIS FINNGTAEIN Hmw3com Hmw2com Hmw4com Hmw1com

64/68 GDLNIIDKKS DAEIQIGGNI SQKEGNLTIS SDKVNITNQI TIKAGVEGGR TIKAGVEGGR TIKKGIDGED TIKAGVDGEN SDKINITKQI SDKINITKQI SDKVNITNQI SOKEGNLTIS SQKEGNLTIS SQKEGNLTIS GSLNITDSNN DAEIQIGGNI GDLNIKNIKA DAEIQIGGNI DTEMQIGGDI GDLNITNEGS Hmw3com Hmw1com Hmw2com Hmw4com

1100 SDSSEAENAN LTIQTKELKL AGDLNISGFN KAEITAKNGS DLTIGNASGG DLTIGNASGG DLTIGNSNDG DLTIGNTNSA LTIQTKELKL AGDLNISGFN KAEITAKNGS KAEITAKDGS TEDLSISGFN KAEITAKDGR TODLNISGFN LTIKTKELKL SSSDATSNAN LTIKTKELKL SDSSEAENAN SDSDATNNAN 1051 Hmw3com Hmw4com Hmw1com Hmw2com

# FIG.101.

					)   
Hmw3com	N. ADAKKVT	N ADAKKVT FDKVKDSKIS TDGHNVTLNS EVKT SNGS SNAGNDNSTG	TDGHNVTLNS	EVKTSNGS	SNAGNDNSTG
Hmw4com	N. ADAKKVT	N ADAKKVT FDKVKDSKIS TDGHNVTLNS EVKT SNGS SNAGNDNSTG	TDGHNVTLNS	EVKTSNGS	SNAGNDNSTG
Hmw1com	D. GTNAKKVT	D.GTNAKKVT FNQVKDSKIS ADGHKVTLHS KVETSGSNNN TEDSSDNNAG	ADGHKVTLHS	KVETSGSNNN	TEDSSDNNAG
Hmw2com	NSGAEAKKVT	VT FNNVKDSKIS ADGHNVTLNS KVKTSSSNGG RESNSDNDTG	ADGHNVTLNS	KVKTSSSNGG	RESNSDNDTG

65	5/6	8		
1200	TGSVEVTAQN /	TGSVEVTAQN	TGNVEIT	NGKASIT
	TKEGTTINAT	TKEGTTINAT	TKTGTTINAT	TTAGSTINAT
	NNNVTSHKTI NISAAAGNVT TKEGTTINAT TGSVEVTAQN	NNNVTSHKTI NISAAAGNVT TKEGTTINAT TGSVEVTAQN	NNNITSHKAV SISATSGEIT TKTGTTINAT TGNVEIT	NKDVTSLKTV NITA.SEKVT TTAGSTINAT NGKASIT
	NNNVTSHKTI	NNNVTSHKTI	NNNITSHKAV	NKDVTSLKTV
1151	LTISAKDVTV	LTISAKDVTV	LTIDAKNVTV	LTITAKNVEV
	Hmw3com	Hmw4com	Hmw1com	Hmw2com

	1201				1250
Hmw3com	GTIKGNITSQ	GTIKGNITSQ NVTVTATENL VTTENAVINA TSGTVNISTK TGDIKGGIES	VTTENAVINA	TSGTVNISTK	TGDIKGGIES
Hmw4com	GTIKGNITSQ	GTIKGNITSQ NVTVTATENL VTTENAVINA TSGTVNISTK TGDIKGGIES	VTTENAVINA	TSGTVNISTK	TGDIKGGIES
Hmw1com	•	AO TGDIKGGIES	•	AOAO	TGDIKGGIES

# FIG. 10.

						66	/68							
T	1300	ISATTGNANI	ISATTGNANI	IKG. TESVTT	•		1350	ADSGKLTSTV	ADSGKLTSTV	ATESLTTQSN	ATVDLTTKSG	1400	NSAKVEAKNG	NSAKVEAKNG
TK		GALTTTAGST	GALTTTAGSŢ	GALTTLAGST	•			NISGNTVTIT	NISGNTVTIT	TISGGTVEVK	TISGNTVSVS		TASTGDLTIG	TASTGDLTIG
•		GQDVTVTADA	GQDVTVTADA	GNTVTVTANS	•			VATGATLAVG	VATGATLAVG	5	9		GTISGNTVNV	GTISGNTVNV
•		GNTLKVSNIT	GNTLKVSNIT	EGALAVSNIS	•			VESSSGSVTL	VESSSGSVTL	•	•		TTSSQSGDIE	TTSSQSGDIE
•	1251	TSGNVNITAS	TSGNVNITAS	SSGSVTLTAT	•		1301	TTKTGDINGK	TTKTGDINGK	SSQSGDIG	GDIS	1351	GSTINGTNSV	GSTINGTNSV
Hmw2com		Hmw3com	Hmw4com	Hmw1com	Hmw2com			Hmw3com	Hmw4com	Hmw1com	Hmw2com		Hmw3com	Hmw4com

Æ,

SKIKATTGEA NVTSATGTIG GTISGNTVNV TANAGDLTVG NGAEINATEG TANAGDLTVG NGAEINATEG GTISGNTVNV NVTSATGTIG SKIEAKSGEA Hmw1com Hmw2com

1450 1401

AATLTAESGK LTTQTGSSIT SSNGQTTLTA KDSSIAGNIN AANVTLNTTG SSNGQTTLTA KDSSIAGNIN AANVTLNTTG SAKGQVNLSA QDSSVAGSIN AANVTLNTTG LTTQTGSSIT LTTEASSHIT AATLTAESGK AATLTTSSGK Hmw3com Hmw4com Hmw1com

1451

STKGQVDLLA QNSSIAGNIN AANVTLNTTG

LTTEAGSSIT

AATLTATGNT

Hmw2com

TLTTTGDSKI NATSGTLTIN AKDAKLDGAA SGDRTVVNAT NASGSGNVTA NATSGILTIN AKDAELNGAA LGNHTWWNAT NANGSGSVIA SGDSTEVNAV NASGSGSVTA NASGSGNVTA NATSGTLTIN AKDAKLDGAA SGDRTVVNAT TLTTVAGSDI KATSGTLTIN AKDAKLNGDA TLTTTGDSKI TLTTVKGSNI Hmw2com Hmw1com Hmw3com Hmw4com

1550 1501

# FIG. 10L.

IQPGVASVEE IQPGVASVEE IOPGIASVDE IQPGVASVEE ISENGRNTVR LRGKEIDVKY ISKNGINTVL LKGVKIDVKY ISKDGRNTVR LRGKEIEVKY ISENGRNTVR LRGKEIDVKY KTSSSVNITG DLNTINGLNI DLNTINGLNI DLITINGLNI DLNTVNGLNI KTSSSVNITG ATSSSVNITG TTSSRVNITG Hmw3com Hmw4com Hmw1com Hmw2com

68/68 VNTQNEFTTK VNTQNEFTTK ALAKLGVSAV RFIEPNNTIT VDTQNEFATR VNTQNEFTTR VIEAKRVLEK VKDLSDEERE TLAKLGVSAV RFVEPNNAIT TLAKLGVSAV RFVEPNNAIT TLAKLGVSAV RFVEPNNTIT VIEAKRVLEK VKDLSDEERE VKDLSDEERE VIEAKRVLEK VKDLSDEERE VIEAKRILEK 1551 Hmw3com Hmw4com Hmw1com Hmw2com

1632 PSSQVTISEG KACFSSGNGA RVCTNVADDG QQ ŏ QP KACFSSGNGA RVCTNVADDG PSSQVIISEG KACFSSGNGA RVCTNVADDG RACFSNSDGA TVCVNIADNG PSSQVTISEG PLSRIVISEG 1601 Hmw3com Hmw4com Hmw1com Hmw2com

#### INTERNATIONAL SEARCH REPORT



In <sup>-</sup>	iticaal	application	No
PCT	7T	2550	

A. CLA								
IPC(5) :A61K 39/02								
US CL According	:424/92 — to International Patent Classification (IPC) or to both	n national classification and IPC						
	LDS SEARCHED		·					
Minimum o	documentation searched (classification system follows	ed by classification symbols)						
U.S. :	424/92; 435/851							
Documenta	tion searched other than minimum documentation to th	ne extent that such documents are included	in the fields searched					
ì	data base consulted during the international search (n q, APS, Biosis, Embase, Scisearch, Chem Abs		, search terms used)					
C. DOC	CUMENTS CONSIDERED TO BE RELEVANT							
Category*	Citation of document, with indication, where a	ppropriate, of the relevant passages	Relevant to claim No.					
Y	Pediatric Infectious Disease Journal 05 May 1990, Barenkamp et al, Bactericidal Activity Following influenzae Acute Otitis Media", 1 337.	, "Development of Serum Nontypable Haemophilus	1-3					
Pediatric Research, Volume 29, No. 4 part 2, issued 1991, Barenkamp S. J., "DNA Sequence Analysis of Genes for Nontypable Haemophilus influenza High Molecular Weight Outer Membrane Proteins which are Targets of Bactericidal Antibody", see page 167A, column 1, abstract no. 985.								
Further documents are listed in the continuation of Box C. See patent family annex.								
Special estegories of cited documents:								
	cument defining the general state of the art which is not considered be part of particular relevance	principle or theory underlying the inve	ention					
·L· doc	rlier document published on or after the international filing date cument which may throw doubts on priority claim(a) or which is	"X" document of particular relevance; the considered novel or cannot be consider when the document is taken alone						
*O* doc	ed to establish the publication date of another citation or other social reason (as specified)  cument referring to an oral disclosure, use, exhibition or other sans	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art						
	cument published prior to the international filing date but later than priority date claimed	*&* document member of the same patent	family					
Date of the 09 MAY	actual completion of the international search	Date of mailing of the international sea	rch report					
Commission Box PCT Washington	nailing address of the ISA/US ner of Patents and Trademarks n. D.C. 20231	Authorized officer  JULIE KRSEK-STAPLES  Telephone No. (702) 208 0106	Warden for					
Facsimile N	o. (703) 305-3230	Telephone No. (703) 308-0196						

Form PCT/ISA/210 (second sheet)(July 1992)★

#### THIS PAGE BLANK (USPTO)